

STIC-Biotech/ChemLib

135240

From: Slobodyansky, Elizabeth
Sent: Friday, October 15, 2004 10:26 PM
To: STIC-Biotech/ChemLib
Subject: 09/712,768

Please search for case 09/712,768:

88

SEQ ID NOs: 1 and 2 against commercial ~~and interference~~ databases.

Thank you.

Elizabeth Slobodyansky, PhD

Primary Examiner

Art Unit 1652
REM 2D83
571-272-0941
MAILBOX 2C70

RECEIVED
OCT 18 2004
TECH/GEN. DIVISION
(STIC)

STAFF USE ONLY

Searcher: _____
Searcher Phone: 2- _____
Date Searcher Picked up: 10/18/04
Date Completed: 10/23/04
Searcher Prep/Rev. Time: _____
Online Time: _____

Type of Search 1
NA Sequence: # _____
AA Sequence: # 1
Structure: # _____
Bibliographic: _____
Litigation: _____
Patent Family: _____
Other: _____

Vendors and cost where applicable

STN: _____
DIALOG: _____
QUESTEL/ORBIT: _____
LEXIS/NEXIS: _____
SEQUENCE SYSTEM: Q1/Q28
WWW/Internet: _____
Other(Specify): _____

This Page Blank (uspto)

GenCore version S.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 22, 2004, 12:24:15 ; Search time 7311 Seconds

(without alignments)
10827.925 Million cell updates/sec

Title: US-09-712-768a-1

Perfect score: 1674

Sequence: 1 atggagagcgccgcatca.....acaagcatccctgcgactaa 1674

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: gb_ba:*
2: gb_ntg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1674	100.0	1674	6 AX113986	Sequence
2	1674	100.0	1674	6 BD015007	Cytochrome
3	1091.4	55.2	2629	1 PDCYCA	Y07533 P.denitrifi
4	1004.8	60.0	2177	1 PDCOX1	X62645 Rhodobacter
5	991	59.2	3211	1 PDCOX2	X05829 P.denitrifi
6	754.4	45.1	286550	1 SMES91785	AL591785 Sinorhizo
7	750.8	44.9	10790	1 AE009044	AE009044 Agrobacte
8	750.2	44.8	11188	1 AE006001	AE006001 Caulobact
9	740.4	44.2	10835	1 AE008010	AE008010 Agrobacte
10	738.2	44.1	10154	1 AE014356	AE014356 Brucella
11	736.6	44.0	11404	1 AE009583	AE009583 Brucella
12	719.6	43.0	346510	1 AP003011	AP003011 Mesorhizo
13	709.2	42.4	4380	1 NWCOXABC	X89566 N.winograd
14	708.8	42.3	1830	1 BJCOXAG	X54318 B.japonicum
15	708.8	42.3	2000	1 BJCOXA	X54800 B.japonicum
16	708.8	42.3	8121	1 BJA242592	AJ242592 Bradyrhiz
17	708.8	42.3	31495	1 BJU33883	U33883 Bradyrhizob
18	708.8	42.3	300900	1 AP005939	AP005939 Bradyrhiz
19	685.8	41.0	349260	1 BX572595	BX572595 Rhodosphe

X74341 R.leguminos
AL646058 Ralstonia
BX640436 Bordetell
BX640451 Bordetell
BX640422 Bordetell
AE012038 Xanthomon
AE004449 Pseudomon
AE016912 Chromobac
AE012504 Xanthomon
AE016774 Pseudomon
AJ389607 Plantago
AJ389606 Plantago
AJ389608 Plantago
AF020565 Ginkgo bi
AF020554 Ehedra v
AE008617 Rickettsi
AY009436 Calycanth
X57692 R.sativus m
AP006444 Brassica
AP006444 Brassica
AF193955 Calycanth
AF193955 Calycanth
AY300014 Brassica
AY300015 Brassica
Continuation (4 of
AF020585 Ephedra d
U77623 Populus tre

ALIGNMENTS

RESULT 1
AX113986
LOCUS AX113986 1674 bp DNA linear PAT 01-MAY-2001
DEFINITION Sequence 1 from Patent EP1103603.
ACCESSION AX113986
VERSION AX113986.1 GI:13940146
KEYWORDS Gluconobacter oxydans
SOURCE Gluconobacter oxydans
ORGANISM Gluconobacter oxydans
Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
Acetobacteraceae; Gluconobacter.
REFERENCE 1
AUTHORS Asakura,A., Hoshino,T. and Shinjoh,M.
TITLE Cytochrome c oxidase complex from gluconobacter oxydans
JOURNAL Patent: EP 1103603-A 1 30-MAY-2001,
F. HOFFMANN-LA ROCHE AG (CH)
FEATURES
source
1..1674
/organism="Gluconobacter oxydans"
/mol_type="unassigned DNA"
/db_xref="taxon:442"
1..1674
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC37869.1"
/db_xref="GI:13940147"
/translation="MADAHGHDEHHEKQFFTRMFMTNHRDKIGLLYLVAAGVVGFI
SVLTFVRLMLDPCGVQVNCLEGLIADASQICTANGHLWNVMVTVGILMFEV
IPALFGGNYLPLQIGAPDVAFPRMNNLSFWLFIAGTAMGVASLFAFGGDLGSG
GVNLVPP:STREAGYSMDLAFVHLGSSSIMGAINMTTFLAMRAPGMLTKVPL
FNSIITLAKLILLALPLVLAGALTMLLDRNEGTTFFNPAGGDDILYQHLLWFSGHP
EYVILLFPGIISRVSTFSKRPVGLPVPYVAVGLGVFWAHMHTVVGSLT
QQSFVLAATVIAVPTGKIFSWIATMGGSVFEPKPMWAFGFMFLFTVGVTGIVL
AQGLDRYDHYVVAHFYVNSIGAFPAIFAGIYFYPKFSGFAPFEWAACKLHFT
FFICANVTFFQHLGRQMPRIYDPEAFALMKNVSYGAPLAFASFLFFIVIFY
TLVAGRRTRPNPWFGEFADLTLEWLPSPPPAHTFTETLPKRSWDKHPSH"

CDS

ORIGIN

Query Match 100.0%; Score 1674; DB 6; Length 1674;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1674; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGGCAGACGCGCCATTCACGGCCATACACCATGAGAGCAAGCTTCTTACGCGC 60
 Db
 1 ATGGCAGACGCGCCATTCACGGCCATACACCATGAGAGCAAGCTTCTTACGCGC 60
 QY 61 TGGTTTCATGTTCGACCAACCAACCAAGACATCGGTCTGTATACCTTTAGCGGCTGGT 120
 Db 61 TGGTTTCATGTTCGACCAACCAACCAAGACATCGGTCTGTATACCTTTAGCGGCTGGT 120
 QY 121 GTTGGTTTCATGTTCGACCAACCAACCAAGACATCGGTCTGTATACCTTTAGCGGCTGGT 180
 Db 121 GTTGGTTTCATGTTCGACCAACCAACCAAGACATCGGTCTGTATACCTTTAGCGGCTGGT 180
 QY 181 GTTCAGTACATGTTCGACCAACCAACCAAGACATCGGTCTGTATACCTTTAGCGGCTGGT 240
 Db 181 GTTCAGTACATGTTCGACCAACCAACCAAGACATCGGTCTGTATACCTTTAGCGGCTGGT 240
 QY 241 CGGACGACACCTTGGAAAGCTCATGGTTACCTTACCAATGATATCTGTATGATGTTT 300
 Db 241 CGGACGACACCTTGGAAAGCTCATGGTTACCTTACCAATGATATCTGTATGATGTTT 300
 QY 301 GTGGGTATCCCGCATTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 360
 Db 301 GTGGGTATCCCGCATTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 360
 QY 361 GCTCCGATATGGCTTCCCGCATTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 420
 Db 361 GCTCCGATATGGCTTCCCGCATTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 420
 QY 421 ACCCGATATGGCTTCCCGCATTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 480
 Db 421 ACCCGATATGGCTTCCCGCATTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 480
 QY 481 GTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 540
 Db 481 GTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 540
 QY 541 GCGATTTCCGGTTCATTTGCGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 600
 Db 541 GCGATTTCCGGTTCATTTGCGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 600
 QY 601 ACACCTTTCTTGAACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
 Db 601 ACACCTTTCTTGAACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
 QY 661 TGGTGGATTTATACAGGCTTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 720
 Db 661 TGGTGGATTTATACAGGCTTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 720
 QY 721 ATCACCATGCTGTCGACCGACCTTCTGCGGACGACCTTCTTCAATCCTGCGCGC 780
 Db 721 ATCACCATGCTGTCGACCGACCTTCTGCGGACGACCTTCTTCAATCCTGCGCGC 780
 QY 781 GGTGACCGGATTTGACCAACATCTGTTCTTTGGGACCGCGAAGTGTACATC 840
 Db 781 GGTGACCGGATTTGACCAACATCTGTTCTTTGGGACCGCGAAGTGTACATC 840
 QY 841 ATCATTCTGCGGCTTTGGCATCATCGGATGTCGTCGACCTTCTCGAAGGCG 900
 Db 841 ATCATTCTGCGGCTTTGGCATCATCGGATGTCGTCGACCTTCTCGAAGGCG 900
 QY 901 GTCTTCGGTACCTGCGGATGTCATGCAATGGTGGCAATCGGTGTTCTGGGCTTGTG 960
 Db 901 GTCTTCGGTACCTGCGGATGTCATGCAATGGTGGCAATCGGTGTTCTGGGCTTGTG 960
 QY 961 GTCTGGGCGCACCATGTGACCGTGGTATGTCGTCGACCGCAATCTTACTTATG 1020
 Db 961 GTCTGGGCGCACCATGTGACCGTGGTATGTCGTCGACCGCAATCTTACTTATG 1020
 QY 1021 CTGGCCACCATGGTATCGCGGTGGCGACCGGCATTAAGATCTTCTGATGATCGCCACG 1080
 Db 1021 CTGGCCACCATGGTATCGCGGTGGCGACCGGCATTAAGATCTTCTGATGATCGCCACG 1080

QY 1081 ATGTGGGGCGGCTCGGTTGAGTTCAAAATCGCGATGCTCTGGGCTTTGGCTTTATGTT 1140
 Db 1081 ATGTGGGGCGGCTCGGTTGAGTTCAAAATCGCGATGCTCTGGGCTTTGGCTTTATGTT 1140
 QY 1141 CTGTTTCAACGTCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1200
 Db 1141 CTGTTTCAACGTCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1200
 QY 1201 TATACGACACCATTTACGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1260
 Db 1201 TATACGACACCATTTACGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1260
 QY 1261 TTTGGATCTTTCGCGGATCTTACCTTTTACATGCGGAAATTTCTGGGCGGCTTTCCG 1320
 Db 1261 TTTGGATCTTTCGCGGATCTTACCTTTTACATGCGGAAATTTCTGGGCGGCTTTCCG 1320
 QY 1321 GAATGGGCTGCAAAAGCTGCATTTCTGGACCTTCTTATCGGTGGAAAGCTCACGTTCT 1380
 Db 1321 GAATGGGCTGCAAAAGCTGCATTTCTGGACCTTCTTATCGGTGGAAAGCTCACGTTCT 1380
 QY 1381 CGGACGACCTTCTTGGGACGTCAGGATGCGGCGGCTTACATCGACTATCCGAAAGCC 1440
 Db 1381 CGGACGACCTTCTTGGGACGTCAGGATGCGGCGGCTTACATCGACTATCCGAAAGCC 1440
 QY 1441 TTGCGGCTGTGGAAACAAAGTCTCGTCTATGTTGGTGGTGGTGGTGGTGGTGGTGGT 1500
 Db 1441 TTGCGGCTGTGGAAACAAAGTCTCGTCTATGTTGGTGGTGGTGGTGGTGGTGGTGGT 1500
 QY 1501 TTTCTTATCGTGGATTTTGTCTATACGTTGCTGCGCGCGCGAGACCGCTCCGAAAC 1560
 Db 1501 TTTCTTATCGTGGATTTTGTCTATACGTTGCTGCGCGCGCGAGACCGCTCCGAAAC 1560
 QY 1561 CCGTGGGCGGAATTCGCGGATACGCTGGAATGAGCTGCGCATCACCGCTTCGCGCCAC 1620
 Db 1561 CCGTGGGCGGAATTCGCGGATACGCTGGAATGAGCTGCGCATCACCGCTTCGCGCCAC 1620
 QY 1621 ACGTTCGAAACGCTGCCAAAGCGCTCGGACTGCGGACAAAGCATCCCTCGCACTAA 1674
 Db 1621 ACGTTCGAAACGCTGCCAAAGCGCTCGGACTGCGGACAAAGCATCCCTCGCACTAA 1674

RESULT 2
 BD015007
 LOCUS
 DEFINITION
 Cytochrome c oxidase complex.
 BD015007
 ACCESSION
 BD015007.1 GI:22555814
 VERSION
 JP 2001169792-A/1
 KEYWORDS
 Gluconobacter oxydans
 ORGANISM
 Gluconobacter oxydans
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
 Acetobacteraceae; Gluconobacter.
 1 (bases 1 to 1674)
 Asakura, A., Hoshino, T. and Araki, M.
 Cytochrome c oxidase complex
 Patent: JP 2001169792-A 1 26-JUN-2001;
 F. HOFMANN LA ROCHE AG
 OS Gluconobacter oxydans
 PN JP 2001169792-A/1
 PD 26-JUN-2001
 PF 17-NOV-2000 JP 2000351502
 PR 17-NOV-1999 EP 99122842.0
 PI AKIRA ASAKURA, TATSUO HOSHINO, MASAKO ARAKI
 PC C12N15/09, C12N1/21, C12N9/08, C12P7/60, (C12N15/09, C12R1:01), PC
 (C12N1/21, C12R1:01), (C12N9/08, C12R1:01), (C12P7/60, C12R1:01), PC
 C12N15/00,
 PC (C12N15/00, C12R1:01)
 CC Cytochrome c oxidase complex
 FH Key Location/Qualifiers
 FT CDS Location/Qualifiers
 1..1674
 /organism="Gluconobacter oxydans"


```

FEATURES
    source
        Location/Qualifiers
            1..2629
                /organism="Paracoccus denitrificans"
                /mol_type="genomic DNA"
                /strain="pD1222"
                /sub_strain="ssp. MR-3"
                /db_xref="taxon:266"
                /clone_lib="EMBL3 lambda"
                233..2544
                    /gene="cyc A"
                233..700
                    /gene="cyc A"
                    /EC_number="1.9.3.1"
                    /note="cytochrome-c oxidase"
                    /codon_start=1
                    /transl_table=11
                    /product="cytochrome c 550"
                    /protein_id="CMA68820.1"
                    /db_xref="GI:45478"
                    /db_xref="GOA:P00096"
                    /db_xref="Swiss-Prot:P00096"
                    /translation="MKISIVATLAAITLALPAAQGDAAKEKFNKCKACHMIQAP
                    DGTDIKGGTGNLVGVGRKIASBEGFKYGEIIEVAEKNPDLTWTADLIEYVD
                    PKWLVMQTDKDKAKTKMTFMKGNQADVVAFIAQNSPDAGDGEAAEGESN"
                233..292
                    /gene="cyc A"
                293..2544
                    /gene="cyc A"
                    /product="cytochrome c 550"
                    /note="cytochrome-c oxidase"
                    /EC_number="1.9.3.1"
                    731..765
                        /gene="cyc A"
                        871..2547
                            /gene="cta DII"
                            871..2547
                                /gene="cta DII"
                                /EC_number="1.9.3.1"
                                /note="cytochrome-c oxidase"
                                /codon_start=1
                                /transl_table=11
                                /product="cytochrome c oxidase beta subunit I"
                                /protein_id="CAA68821.1"
                                /db_xref="GI:45479"
                                /db_xref="GOA:P98002"
                                /db_xref="Swiss-Prot:P98002"
                                /translation="WADAHVGHGDHDTGRGFTFWFSTNKHDIILYLFAGIVGL
                                ISVCFIVYRMELQHPGVQVYMCLEGARLIADASCTPNGLWNVMTYHGVLMWFFV
                                VIPALFGGGNTFMFLHGAPMDAPRLNLSYMYVCGVALVSLIAPGQNDQMS
                                GVGWVLIPLSTIAGSYMDLAFIAVHVSAGSILGAINIITPLNRAPOMTLPKVP
                                LEAWSVFTAMILLISLPALAGATIMLLMDNRFQTFDPAGGDPVLYOHLIFPEGH
                                PVYIILPFGIISVLTSTPAKPIFGVLPWLVAIAIGILGVVWAHMYTAGMSL
                                TQAVEMLATMTIAPVTKGVSWIATWNGSIEFKTMLNAPFLFLFTVGGVTGVV
                                LSQALDRYHDTYVVAHFHYVMSLGAVFGIFAGVYIWMKSGRQYEPWAGQLHF
                                NMFIQSNLIFFPQHLFGKQGRPRVIDYVDFVYNNISSIGAVISFASFLFIQIVF
                                YTLFAGKRVNPNYVNEHADLTLEWTLPSPPPEHTPETLPKREDWDRAHAH"
                                2586..2615
                                    /terminator
                                ORIGIN
                                    Query Match
                                    Best Local Similarity 85.2%; Score 1091.4; DB 1; Length 2629;
                                    Matches 1318; Conservative 0; Mismatches 356; Indels 3; Gaps 1;
                                1 ATGGCAGCGCGCCATTACGGCCA---TGACCAACCATGAGAACCAAGGCTTCTTACG 57
                                871 ATGGCAGACGCGCGCTTACCGCCACCGGTGACCATCATGACACCGCGGCTTCTTACC 930
                                58 CGCTGTTTCATGTCACCAACACACACATCGGTCTGCTATACCTTGTAGCGGCTGGT 117
                                931 CGCTGTTTCATGTCACCAACACACACATCGGTATCCTTTACCTGTTTACCGCGCGG 990
                                118 GTTGTGTTTTCATTCGGTCTGTTACCGCTTACATGCGGCTTGAGCTGATGATCCG 177

```

```

Db
991 ATCTCGTGGCTGATCTCGTATGCTTACGGTCTATATATGCGGATGGAACATCGAGATCCG 1050
QY
178 GGTGTTAGTACATGTGCTTGAAGCGCACGTCTGTATGCGGGATGCCCTCGCAGACATGT 237
Db
1051 GCGGTGCAATACATGTGCTGGAAGCGCGGTCTCATCGCCGACGCTCGCGGAATGC 1110
QY
238 AGCGGGAACGACACCTCTGGAACGTATGTTTACCTACCATGATGTTCTCATGATGTTT 297
Db
1111 ACCCGAAGGACACCTCTGGAACGTATGATACCTACCAACGCGGTCTCATGATGTTT 1170
QY
298 TTGTGGGTATCCCGCATTTGTTGGTGGTATTTGGTAACTATCTCATGCCCTGCAAAATC 357
Db
1171 TTCTGCTGATCCCGGCACTGTTCGGCGTTTCGGCAACTATTTTCATGCCGCTGCATATC 1230
QY
358 GCGCTCCGGATATGGCTTCCCGGTATGAACACCTGTGCTGCTGCTGCTGCTGCTGCTGCC 417
Db
1231 GCGCCCCGGACATGGCCCTTCCCGCGGTGAACAACTCTCTACTGATGATGTTCTGTC 1290
QY
418 GGTACCGCGATGCGCGTGGCTTCTGTTCCACACCGCGCGGTGACGCTCAGCTGGGTTCC 477
Db
1291 GCGTGGCCCTGGGCTGCGCTCTGCTGCGCGCGGCGGCAACGACAGATGGGTTCC 1350
QY
478 GCGTGTGGTGGTCTGTACCCCGCGCTGTGCAACCGCGGAGCTGCTATTCGATCGAC 537
Db
1351 GCGGTGCGCTGGGTGCTCTACCCCGCTCTCGACCAACCGAGCGGCTTATTCATGGAC 1410
QY
538 CTCGCGATTTTCGCGGTTCACTTGTGCGGTGCTCTCTCGATCATGGCGGAGTCAACATG 597
Db
1411 CTGGGATCTTTCGCTCCAGCTCTCGGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1470
QY
598 ATACGACCTTCTTGAACATGCGCGCCCCCGCGATGACGCTGCACAAAGTCCGCTTCTTC 657
Db
1471 ATACCACTTCTCTCAACATGCGCGCACCGGCGATGACGCTGTTCAAGGTGCGCTGTTT 1530
QY
658 TCGTGTGCTGATCTTATCAGGCTTGGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 717
Db
1531 GCTTGGTGGTCTTATCAGCGCTGGCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1590
QY
718 GCAATCACTGCTGTGACCGACCTTAATCTCGGACGACCTTCTTCAATCTCTGCTGGC 777
Db
1591 GCGATCACTGCTGCTGTATGACCGACCTTCTCGGACGACCTTCTTTCGATCCGCGCGC 1650
QY
778 GCGGTGACCGGATCTGTGACCAACATCTCTGCTGCTTCTTGGGACCCGGAAGTGTAC 837
Db
1651 GCGGCGACCGGCTGCTTACAGACATCTCTGCTGCTTCTTGGGACCTCCGAGGCTTAT 1710
QY
838 ATCATCATCTTGGCCGCTTGGCATCATCAGCATGTGCTGCTGACCTTCTCGAAAAAG 897
Db
1711 ATCATCATCTTGGCGGCTTGGCATCATCAGCATCATCAGCATCTCTGACCTTCCGCAAG 1770
QY
898 CCGGTCTTGGTTACCTGCGGATGCTATGCAATGGTGGCAATCGGTGTTCTGGGCTTT 957
Db
1771 CCGATCTTGGCTACCTGCGGATGCTGCTGGCATGCTGCGGATCGGCTATCTGCGGCTTC 1830
QY
958 GTGCTGTGGCGCACCATATGATACCGTTGGTATGCTGCTGACCCAGCAATCTCTACTTC 1017
Db
1831 GTGCTGTGGCGCACCATATGATACCGCGGATGCTGCTGACCCAGCGGCTTATTTTC 1890
QY
1018 ATGCTGGCACCATGATGATGCGGTCGCGACCGGATTAAGATCTTCTGCTGGATCGCC 1077
Db
1891 ATGCTGGCACCATGATGATGCGGTCGCGTCGCGATCCGCGATCAAGGCTTCTTCTGCTGG 1950
QY
1078 ACCATGTGGCGGCTGCTGATGATTAATCGCGATGCTCTGCGGCTTCTGCTGCTTATG 1137
Db
1951 ACCATGTGGCGGCTGATGATTAAGATGCGGATGCTCTGCGGCTTCTGCTGCTTCTG 2010
QY
1138 TTCTGTTTACCGTGGTGTGTGACCGGATGCTGCTGCGGCTTCTGCGGCTTCTGCGGCT 1197
Db
2011 TTCTGTTTACCGTGGTGTGTGACCGGCTGCTGCGGCTTCTGCGGCTTCTGCGGCTTCTG 2070
QY
1198 GCATATCAGCACCTTATACCGTGGGCTTCTTCAATGATGATGCTGCTGCTGCTGCTG 1257
Db
2071 GTCTATCAGCACCTTATACCGTGGGCTTCTTCAATGATGATGCTGCTGCTGCTGCTG 2130

```


Db 1048 TGGTCTTCGCGCCACCGGAGGCTACATCATCTGCTGCGCGCTTCGCGCATGCTCAGC 1107
Qy 871 CATGTCGTGTGACCTTCTCGAAAGCCGCTCTTCGTTTACCTGCGGATGGTCTATGCA 930
Db 1108 CACGTCATCGGACCTTCGCCAAGAGCGATCTTCGGCTATTTGCCATGGTCTATGGC 1167
Qy 931 ATGTGGCAATCGGTCTTCGCGCTTGTGCTCTGCGCGGACCAACATGATACACGTTGGT 990
Db 1168 ATGTGGCAATCGGCTGCTGCGCTTGTGCTCTGCGCGGACCAACATGATACACGCGGC 1227
Qy 991 ATGTGCTGACCCAGCAATCTACTCTCATGCTGCGCACCATGCTGATCGCGGTGCGGAC 1050
Db 1228 CTGAGCCTCACCGAGAGCTACTTTATGATGCGGACCATGCTGATGCGAGTGCACACC 1287
Qy 1051 GGCATTAAAGATCTTCTCGTGATCGCCACGATGTGGCGGCTCGGTGAGTTCAATCG 1110
Db 1288 GGCATCAAGATCTTCTCGTATCGGACCATGCGGCGGCTCGATCGAGCTCAAGACG 1347
Qy 1111 CCGATGCTCGGCGCTTTCGCTTTATGCTCTGCTGCTGCGGTGCTGATGCGCGGTATC 1170
Db 1348 CCGATGCTCGGCGCTTTCGCTTTCTCTCTTACCGTGGCGCGCTCAGCGCATC 1407
Qy 1171 GTGCTGCCAAGCGGTCTGCGCGCTGCTATCATGACGACCATTAACGTTGCGGCGCAC 1230
Db 1408 GTGCTGAGCGGCGGAGCGCTGCGCGCTATATACGACGACCTACTATGCTGCGGCGCAC 1467
Qy 1231 TTCCATTATGATGCTGCGGTGCGATCTTTCGATCTTCGCGGTATCTACTTTTAC 1290
Db 1468 TTCCATTATGATGACCTCGGCGGCTTTCGCGATCTTCGCGGATCTACTTCTG 1527
Qy 1291 ATGCCGAAGTCTCGGCGCGCTTTCGCGAATGGCTGCTCAAGCTGCACTTCTCGGACC 1350
Db 1528 ATCGGCAAGATGTCGGCGCGGCAATATCGGAATGGCGGAGCTGCTATTTCTGGATG 1587
Qy 1351 TTCTTCATCGGTGCGAAGCTGCTGCTTTCGCGGAGCTTCTCGGAGCTCAGGATG 1410
Db 1588 ATGTTCGTGGCGGAGCACTTTCGCGGAGCTTTCGCGGAGCTTTCGCGGCGGAGCATG 1647
Qy 1411 CCGCGCGTTTACATCGACTATCGGAGCTTTCGCGGAGCTTTCGCGGAGCTTTCGCGGAG 1470
Db 1648 CCGCGCGCTTACATCGACTATCGGAGCTTTCGCGGAGCTTTCGCGGAGCTTTCGCGGAG 1707
Qy 1471 GTGCGGTTCGCGCTTTCGCTTCTGCTTCTTCTGCTGATCTTCTGATGCTG 1530
Db 1708 GCGCGCTTCTGCTTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1767
Qy 1531 GTTGTGCGCGCGGCGGAGCGCTTTCGCGGAGCTTTCGCGGAGCTTTCGCGGAGCTTTCG 1590
Db 1768 ACGGCGCGCGGCGGAGCGCTTTCGCGGAGCTTTCGCGGAGCTTTCGCGGAGCTTTCG 1827
Qy 1591 TGGACGCTGCATCAGCGCTTTCGCGGAGCTTTCGCGGAGCTTTCGCGGAGCTTTCGCGGAG 1650
Db 1828 TGGACGCTGCATCAGCGCTTTCGCGGAGCTTTCGCGGAGCTTTCGCGGAGCTTTCGCGGAG 1887
Qy 1651 TGGACGCTGCATCAGCGCTTTCGCGGAGCTTTCGCGGAGCTTTCGCGGAGCTTTCGCGGAG 1674
Db 1888 TGGACGCTGCATCAGCGCTTTCGCGGAGCTTTCGCGGAGCTTTCGCGGAGCTTTCGCGGAG 1911

RESULT 5
PDCOX2
LOCUS
DEFINITION P.denitrificans COI gene.
ACCESSION X05829
VERSION X05829.1
KEYWORDS Gi:994806
unidentified reading frame.
SOURCE
ORGANISM Paracoccus denitrificans
Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
Rhodobacteraceae; Paracoccus.
REFERENCE
AUTHORS Raitio,M., Jallio,T. and Saraste,M.

TITLE Isolation and of the genes for cytochrome c oxidase in Paracoccus denitrificans
JOURNAL EMBO J. 6, 2825-2833 (1987)
REFERENCE 2
AUTHORS Saraste,M.
TITLE Direct Submission
JOURNAL Submitted (04-JAN-1991) Saraste M., EMBL, Meyerhofstrasse 1, Heidelberg, Germany
REMARK revised by [3] MAT
REFERENCE 3 (bases 1 to 3211)
AUTHORS Saraste,M.
TITLE Direct Submission
JOURNAL Submitted (20-SEP-1995) Saraste M., EMBL, Meyerhofstrasse 1, Heidelberg, Germany
COMMENT On Sep 21, 1995 this sequence version replaced gi:45474.
FEATURES
source
1..3211
/organism="Paracoccus denitrificans"
/mol_type="genomic DNA"
/db_xref="taxon:266"
/map="cox2 locus"
218..222
RBS
229..564
CDS
/note="unnamed protein product; ORF4"
/codon_start=1
/transl_table=11
/protein_id="CAA29273.1"
/db_xref="GI:45475"
/db_xref="GOA:P08304"
/db_xref="Swiss-Prot:P08304"
/translation="MADISMPKASALGETSRAGASIDDAFLALSKVVFADGALDKRT
KQLIAYAVAHVTQCPWCIEGHVKGARBGATNEQIMEAIWVAEMRGAAYAHANKAL
AVLDEIDGA"
mat_peptide 229..561
/product="ORF4"
repeat_unit 628..640
/note="inverted repeat A"
TATA_signal 686..689
repeat_unit 713..725
/note="inverted repeat A"
RBS 765..769
gene 777..2444
mat_peptide 777..2444
/gene="COI"
/gene="COI"
/product="cytochrome-c oxidase"
/EC_number="1.9.3.1"
CDS 777..2441
/gene="COI"
/EC_number="1.9.3.1"
/codon_start=1
/transl_table=11
/product="cytochrome-c oxidase"
/protein_id="CAA29274.1"
/db_xref="GI:995678"
/db_xref="GOA:P08305"
/db_xref="Swiss-Prot:P08305"
/translation="MSAQISDSIEKKRGFTTFMSTNKHDKIGLVLYFTAGLAGLTSV
TLTVYRMELOHPGVCMCLGEMLVADAACECTPNALWNVVTVYTGILWPFVVP
ALFGGCTFMPFLHIGAPDPAFRLNLSYWLVCVSLASLSLSPGSDQPCAGVG
WVLYPSTTEAGAMDALFAVHVSATGSLGAINITITFLNMRAPGTLTKVPLFA
WAVITAMWLLSLPLAGGITMLMDRNFQFFDPAGGDPVLYQHILWFFGHEV
YMLILPGFGIISHVISTFARKPIFGYLPMLAAIAFLGFIWAHMYTAGMSLQQ
TFQMATIATVPTGKVFSTATMGSGIEFKTLMWALFLFTVGGTVGVVIAQGS
LDRVYHDTYIVAHFVYVMSLGAIPAI FAGTYVWIGKMSGROYPEWAGOLHFMFMFIG
SNLIFPQHLPGQGMERVIDYVPEFSYNNISSIGAVISFASLFFIGIVFYTLFA
GKPVNPNYNNHADTLEWLPSPPEHTFEILPXPEDWDRAQHR"
2744..2788
terminator
repeat_unit 2744..2755
/note="imperfect inverted repeat B"
repeat_unit 2764..2775
/note="imperfect inverted repeat B"
ORIGIN

Laboratoire de Biologie Moléculaire des Relations
Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,
France, Laboratoire de Génétique et Développement UMR6061-CNRS,
Faculté de Médecine, 2 avenue du Pr. Léon Bernard, F-35043 Rennes,
France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,
Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,
D-33615 Bielefeld, Germany, Unite de Biochimie physiologique,
Université Catholique de Louvain, Place Croix du Sud 2, Bte 20,
B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculté
des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6,
B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzou@toulouse.inra.fr
<http://sequence.toulouse.inra.fr/meliloti.html>.

FEATURES

```

source
1..286550
/organism="Sinorhizobium meliloti"
/mol_type="genomic DNA"
/strain="1021"
/db_xref="taxon:382"
70..813
/gene="gst11 OR SMC00916"
70..813
/gene="gst11 OR SMC00916"
/EC_number="2.5.1.18"
/functions="miscellaneous"
/notes="Product confidence : putative
Gene name confidence : putative
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE GLUTATHIONE S-TRANSFERASE PROTEIN"
/protein_id="CAC45361.1"
/db_xref="GI:15073720"
/db_xref="GOA:Q92RR5"
/db_xref="TrEMBL:Q92RR5"
/translations="MADLSFFPIIARWPAKNPDIILQSLTPNGVKVSAIEELGLA
YEAFISFDNEQSPFSLNPNRIPAIIDNPGDGPFGIFGFSGALLIYLAQKTG
KLIPAAARYETLCWGMFGFMFGFQGFHFFFAADKVANNPVPVRYDRAKR
LLGVLEARLEGROWLGMGRNTPSPTSRPIPGWKGSSVAPRCSTIRASRNSWTGI
AAWHALRPREARSRSRARSRLPAGD"
588..779
/gene="SMC00915"
588..779
/gene="SMC00915"
/functions="miscellaneous; hypothetical/partial homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC45362.1"
/db_xref="GI:15073721"
/db_xref="TrEMBL:Q92KJ5"
/translations="MADGGEYTIADIATYPWVEGARPYGGAELVDYKSPFNVDWV
DRGLAPAAQKGMEIPRKE"
867..2084
/gene="SMC00914"
867..2084
/gene="SMC00914"
/functions="small molecule metabolism"
/notes="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental

```

```

/transl_table=11
/product="PUTATIVE OXIDOREDUCTASE PROTEIN"
/protein_id="CAC45363.1"
/db_xref="GI:15073722"
/db_xref="GOA:Q92RR4"
/db_xref="TrEMBL:Q92RR4"
/translations="MSGRLLVVVGGOQAALFVAKLRKQMRPTTVVAERASLPYQRP
PLSKYLREMTLRLLLYRPAWAEHEIDIRLSTTVTRDLAKQVALSDGSMITVE
TLAFATGTPRLPRAVGGDLAGVVFVRDFRADRLABEMQGRVILVVGGYIGLEA
AAVARTSGLEVTIEMADRIILQVSAATSATVREIHRSHGVDIRERTGLHRLICDNG
RVTAELSDGSGVIPVDIVIVIGVAANDALAEAGIETANGIVVDVSHGRTSDPTIVAM
GCAVLPWDGMRIRLESQNAVDQAEAAVAAGLAGTDPDYPDKPMFSDQYDKLQIAG
FGLGHDTLVROGQSGSVVYFRQGLIAVDINDAKAVTVTGKKLLEAGATPDRLL
LADQVDLKALLV"
Complement (2114..2572)
/notes="Sm-2 OR SMC04876
REPEAT SM-2
predicted by Homology"
/evidence=not_experimental
/complement (2585..4222)
/gene="groEL1 OR SMC00913"
/complement (2585..4222)
/gene="groEL1 OR SMC00913"
/function="cell processes; chaperoning"
/notes="predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="60 KD CHAPERONIN A PROTEIN"
/protein_id="CAC45364.1"
/db_xref="GI:15073723"
/db_xref="GOA:P35469"
/db_xref="Swiss-Prot:P35469"
/translations="MAAEKVFGRSAREKMLRGVDILADAVKVTLPKGRNVIDKSF
GAPRTIKDGSVAKEIELEDFENMGQMYREASKNTDIAGDGTATVLAQAIARE
GAKAAVMGPMDLKRGIDIAVEVWDLAKAKINTIDEVAQVGTISANGKQIGL
DIAMOKGVNEGVIITVEEAKTAELEVVGEVQFDRGYSFPVTNPKWADLEDA
FILLKELSNLQALPVLBAVCTGKPLIIAEDVEGEALATLVNKLKRGGLKIAAV
KAPFGRRKMLIEDAILTGGTVISDLGKLESVTLMLGRKAKVSIITKENTTVVD
CAGKSDIERVAQIKAEIETSDYDRKQLERLAKAGVAVIRVGGATEVEVREK
KDRIDALNATRAAVQEGIVPGGVALLRSRVKIVKGENDDQDQGVNIIVRRALQSPA
KQIVENAGDEASIVVQKILEKNTDDFGYNAQTGEYGDMIAMGIIDPVKVVRTALQDAA
SVASLITTEAMTAELPKDAPAMPGGMGGMGMDMM"
Complement (4298..4594)
/gene="groES1 OR SMC00912"
/complement (4298..4594)
/gene="groES1 OR SMC00912"
/function="cell processes; chaperoning"
/notes="predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="10 KD CHAPERONIN A PROTEIN"
/protein_id="CAC45365.1"
/db_xref="GI:15073724"
/db_xref="GOA:P35473"
/db_xref="Swiss-Prot:P35473"
/translations="MASTNFRPLHDRVVRVRRVSEETKGGIIPDTAKEKPOEGEIV
AVGSGARDESGKVPLDVRAGDRILFKKSGTEVKINGEDLLIMKADINGVIG"
4863..5690
/gene="SMC00911"
4863..5690
/gene="SMC00911"
/functions="miscellaneous; hypothetical/global homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Framed"
/codon_start=1

```

[illegible]

193	Qy	TGCTTTGAAGGCGCACGCTCTGATCG- - -CGATGCTCTCGACACATGTATACGGCGAATCGG	248
2159	Db	CACGCTCTGGCTTCCATGCTCTACGGCTCGAGGGGATGCCCATCGACGCGGTAAG	2218
249	Qy	ACACCTGTGGAACGTCATGGTTACCTAACAATGGTATTCTGATGATGTTCTTTGTGGGTAT	308
2219	Db	-CACATGTTCAACGTTTCAACACACGCGCACGGCTGATCATGATCTTCTTCATGGTTCAT	2277
309	Qy	CCCGCATTTGTCGGTGGTTTTGGTAACTATCTGATGCCGTGCAAAATCGCGCGTCCCGA	368
2278	Db	GCCGGGATGATCGGCGGTTTTGGCAATGAGATCCGATCATGATCGGCGGCTCGA	2337
369	Qy	TATGGCTTCCCGCGTATGAACAACTGCTGCTTTGGCTTTTCAATGCGCGGTACCGGAT	428
2338	Db	CATGGCTTCCCGCGCTCAACAATATTTCCCTTCTGGCTCATCGTTCCGGCTTTCATCCT	2397
429	Qy	GGGGCGGCTTTCGCTGTTGCACACGGCGGTGACGGTCAGCTGAGTGGTTCGGCGGTGGTG	488
2398	Db	GCTGCTCTCTGCTCTTTCGTGGAAGTCCGGCGGCTGCTATGCGTGGTGGCGGCTG	2457
489	Qy	GGTTCTGTACCCCGCTGTCGACC- - -CGCGAAGCTGGCTATTTCGATGGACCTCGCGAT	545
2458	Db	GACATGTATTCGGCTCTGTGCACAGTGGCATGCCGGCGCGGTGGATCTTGGAT	2517
546	Qy	TTTCGGGGTTCATTTGTGCGGTGCTCTCTGATCATGGGCGGATCAACATGATCACGAC	605
2518	Db	CTTCTGCTCAGTGCCTCGGTGCACTCTCGATCTCTCGCGCCATCAACTTTCATCACAC	2577
606	Qy	CTTCTTGAACATCGCGGCCCGCGGATGACGCTGCACAAAGTCCGCTGTTCTCGTGTC	665
2578	Db	CATCTGAACATCGCGCGCGCGGCAAGACGTGCACAAAGATGCCGCTGTTTGGCTGGTC	2637
666	Qy	GATCTTTATCACGGCTTGGGTGATCTGTGCGGTGCGGTTCTGGCTGGTGCATCAC	725
2638	Db	GGTCTGGTCACCGCATTCCTGCTCGGTTTCGTTCCGTTTGGCAGGTGGCATCAC	2697
726	Qy	CATGCTGTGACCGACCGTAACCTTCGCGACGACCTTCTTCAATCTGCTGGCGCGGTGA	785
2698	Db	CATGCTTCTGACCGACCGTAACCTTCGCGACGCGGTTCTTCTCGCGAGGCGCGGTGA	2755
786	Qy	CCCGATTCTGTACCAACACATCTCTGTGGTTCTTTGGCACCCGGAAGTGTACATCAT	845
2758	Db	TCCGATCCTCTACAGCATCTGTTCTGGTTCTTCGGTTCACCGGAAGTGTACATCTCAT	2817
846	Qy	TCTGCCGGCTTTGGCATCATACGCCATGTGCTGTGCACCTTCTCGAATAAGCCGGTCTT	905
2818	Db	CCTGCCGGTTTGGCATCATACGCCATGTGCTGTGCACCTTCTCGAATAAGCCGGTCTT	2877
906	Qy	CGGTTACCTGCCGATGCTTATGCAATGGTGGCAATCGGTGTTCTGGGCTTTGTGCTGTG	965
2878	Db	CGGTTATCTCGCATGGCTTACGCCATGTCGCCATCGTGCGGTTCATCGTCTG	2937
966	Qy	GGCGCACCATGTACACCGTTGGTATGTGCGTGAACCCAGCAATCTTCTCATGCTGGC	1025
2938	Db	GGCCACCATGTACACCGTTCGGCTGTTCGCTCGAGGCACACGGCTACTTCGTTTTCGC	2997
1026	Qy	CACCATGTGATCGGGTCCGACCGCATTAAGATCTTCTGTGATCGGCACCATGTG	1083
2998	Db	CACCATGTGATCGGGTCCGACCGGATCAAGATCTTCTGTGATCGGCACCATGTG	3057
1086	Qy	GGCGGCTCTGGTTGAGTTCAAAATCGCGATGCTCTGGGCTTTGGGCTTTATGTTCTGTT	1147
3058	Db	GGGCGGTTCTGCTGACCTTCTGCACCCGATGGTCTGGCAATCGGCTTTATCTTCTGTT	3117
1146	Qy	CAACCGTGGGTGTGACCGGTATCTGTGGCCCAAGCGGTCCTGGACCGTGCATATCA	1207
3118	Db	CACGGTCTGGTGTCAACCGGTGTGCAGCTTCGCAATGCGCGTCTCGACCGTTCGCTGCA	3177
1206	Qy	CGACACCTATTACGTGGTGGCGACTTCCATTAATGTGATCTCGCTGGGTGCGATCTTTCG	1265
3178	Db	CGACACCTATTACGTGGTGGCTCACTTCCACTACGTGCTGCTCGCTCGGCGCGTCTTTCG	3237
1266	Qy	GATCTTCGCGGTATCTACTTTTTATCATGCCGAAGTCTCGGCGCGGCTTTCGCCGAATG	1327

VOGAEAKIVVSGPKSYVDRLLKVEGGRISLSDINLADGALMTLPGDLQVRSRDRDKMI
TVAPDVTREVEVSGDLOTRRYEQALNLDVSGSDVEAGKVTETVTVENSNGSYAD
LGLKAEAALEIVSGSGAVPATGKAKIDISGSDVELRTEPKSVTSEITGSGEVR

gene

4964..6436

CDS

/gene="CC3411"

/note="CC3411"

/note="identified by match to PFAM protein family HMM

PF01523"

/codon_start=1

/transl_table=11

/product="13D protein"

/protein_id="AAK25373.1"

/db_xref="GI:13425123"

/translation="MSAFGLTCRTMNAPI PASSLSILDAAGVDPDRARAILGSPALA
GADGELFLERSSEAFVDFDGLKASYSDSGEFLRVAGETAGYAHAAEISPAAI
RRADSASAKRHAGVSAEGPRATNEKLYGSDNPVSAPDFDKVALLQEIADAFAR
DRPVOMASACERDVEILRADGLRIDVRPLRVNVQVTVKDGRRSSGAGG
RAGFAVSPDKQDOVDLEALMALNLDVACPAGENDVVLPGWNGVILHEAVGCH
LEGDFNRKGISASGSGIGERVAAPCVTVFDDSGISPGRRSLVDDEGPTERTILIED
GILVNGHDMRSLRMGMKATGNRRQSTAHMPMRMTNIMLAGEHDPQEMIASLKR
GLYCAFGGGQRLITNGKVFQCTEAYLVDEDKITAPVKGATLIGDGPALTRVTMIG
NDFDFDGIQVCGKSGGVPGVGVGPSLKITGLTVGGTAV"

gene

/gene="CC3412"

complement(6601..7098)

/note="CC3412"

/note="identified by match to protein family HMM"

/codon_start=1

/transl_table=11

/product="acetyltransferase, GNAT family"

/protein_id="AAK25374.1"

/db_xref="GI:13425124"

/translation="MSRLSIPILADLNADNPAGALNNAHAVELSMLEPERLAHV
AQFLARRGVADALLTDEHDADYSPNPLFRERYPPVYVDRVAVDASRGELA
RALYDDLDAKAAAGHARIVCEVNSDPNPVSDAFHALLGVFGVGTAEIHGSKIVTY
LERRL"

gene

complement(7386..9266)

CDS

/gene="CC3413"

complement(7386..9266)

/note="CC3413"

/note="identified by match to protein family HMM"

/codon_start=1

/transl_table=11

/product="peptide transporter"

/protein_id="AAK25375.1"

/db_xref="GI:13425125"

/translation="MNIVAVGILVTLTGIPVLLQLLKHPRGLIICFLAEMBRFS
YGMRAILFYLTOHFLPDKAASGHVSTALVYLPLIGGLADKLGRKAVAPG
AVLLVAGHAWALEGEKAVQCTLYQCATVEQAEQSGONEAKLVUDGKAYEATKV
GDEIKGLPANAPLPAVLKADYKLDVAGRDLYLNIWFWSLSLILGFLKPNIST
IVQGLPQGRPRDPGFTLYGINUGSWASILCGLLVNVVWAGFLAGVGMLAG
FVFLGKPMIDKAPPEPKVLEKRVAGPINRELAISLVGLVIGLFFLVQFNALV
GVANTGILASLGYIGFFMFKCAKEERERLALAVELIFGAVFWTLFEQAGSSLSLF
AANTVQDLVREPLRIFSGAVTLATPDQAKAGIDAASTFWVDTSFNAPQTAQNAWG
ILIFAPVFAVMTFLQAGDPNPVKFGLALIQVAGFLQLQWGSTADGAPKPLI
FLVMVMLHTTGELCKSPVGLSOMTKLSPASVSEFMAVFLAVIAIAOWVGGKLAGLA
ATEVGGQVLDPGMALAASLKVENIIGLVAVALIGIFLLSPVLKXWSHADNVGAPK
XA"

gene

complement(9726..11087)

CDS

/gene="CC3414"

complement(9726..11087)

/note="CC3414"

/note="similar to GP:3688156; identified by sequence

similarity; putative"

/codon_start=1

/transl_table=11

/product="glutamate-cysteine ligase"

/protein_id="AAK25376.1"

/db_xref="GI:13425126"

/translation="MADTASVQERQLTDLDTAYFAKSGKPEAFRVGAHEKFGFYL
GSEHPVYEGDKGVALHATLQRFQWTPVMEGPTIIGLRNGANVSLRPGQFELSGA

PLSMHEICBETGQHLDEVKTVADELGLGFTGLGFSFKWTRQGVIMPGRYVIMRNY
MPKVNGLIGDMLLRTCTVQANLDFSEADMVAKFMSLALQPIATLAFNSPFTBQKP
NGFLSARANVWTDTPENRTGLDFVREGDFDPERVAVLDVPMYFVKGDKYIDVAG
RSFRDLECKLPELGEIATMDADHTTAPFVRLKTYLEMGRADGAPNSRLCALP
ALMTGVYDDAALAAWDLCKDWTLEDRGLRNVFKGLKAQVAGADQVAKDFVFA
VAKAGLNRAQLNGGLDETIIYGLLEQLADSGITPAEKLILSYNGAWGGDIERVFTD
CAY"

ORIGIN

	Query Match	44.8%	Score 750.2	DB 1	Length 11188
	Best Local Similarity	67.5%	Pred. No. 1.4e-145		
	Matches 1122	Conservative 0	Mismatches 518	Indels 21	Gaps 4
QY	4	GCAGAGCCGCCATTCACGGCCATGACACCACTGATGAGACCAAGGCTTCTTCAAGCGCTGG	63		
DB	1740	GCAGACACTCACGACACGCCCCACGACGACCAACAGCCGCCCTTCTTCCGCTCGCTGG	1681		
QY	64	TTCAATGTCGACCAACACAAAGACATCGCTCTGCTATACCTTGTAGCGGCTGGTGTGT	123		
DB	1680	TTCTTCTCCAGCAACACAAAGACATCGGACGCTGTACATCTCTGTCGCCATCATGGG	1621		
QY	124	GGTTTCATTTCCGCTCTGTTTACCGCTCTACGCTGACGCGCTTGAAGTATGGATCGGGTGT	183		
DB	1620	GGCATCGTCGCTGGCGCGCTTTTCGGGCTCATCCGCTGGGAGCTGATGCGCGCGGCAAT	1561		
QY	184	CAGTACATGTGCTTGAAGGCGCACGCTCTGATCGCGGATGCTCGCAGACATGTACGGG	243		
DB	1560	CAGGTGTTCTCCACACCGG-----CATGCTGCCACGATGGCATCTTCAAG	1513		
QY	244	AACGGACACTGTGGAAAGTATGTTTACCTACCATGATGTTTCTGATGATGTTCTTTGG	303		
DB	1512	GGCAAGCACGCGCTATAACGCGCTGTCACGCGCCACGCGCTGATCATGATCTTCTCATG	1453		
QY	304	GGTATCCCGCATTTGTCGCTGTTTGGTAACTATCTGATCGCGCTGCAATCGCGCT	363		
DB	1452	GTCACTGCTGGATGATCGCGGTTTCGCAACTGTTCTGTTCCGATCATGATCGCGG	1393		
QY	364	CCGATATGCGCTTCCCGGATGAAACCTGCTGCTTCTGCTGCTGTTCTTCCGCTGATCC	423		
DB	1392	CCGACATGCGCTTCCCGGATGAAACATCTGTTCTGCTGCTGATCGCGCTG	1333		
QY	424	CGCATGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	483		
DB	1332	ATCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1276		
QY	484	GGTTGGTCTGTACCGCGCGCTGTCGACCGCGG---AAGCTGGCTATTCGATGACCTC	540		
DB	1275	GGTTGACGATCTATCCGCGCTGTCGACCATCGGTCAAGGGTCCGCGCATGACCTG	1216		
QY	541	GGATTTTCGCGCTTCACTTGTGCGGTGCTCTCTCGATCATGGCGCGATCAACATGATC	600		
DB	1215	CGCATCTGCTGCTCACCTGCGCGGCTGCTGATCTCTGGCGCGATCACTTCATC	1156		
QY	601	ACGACCTTCTTGAACATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	660		
DB	1155	ACCACGATCTTCAATATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1096		
QY	661	TGCTGATCTTATCAGGCTTGGCTGATCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	720		
DB	1095	TGCTGCTGCTGATCACCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1036		
QY	721	ATCACATGCTGTGACCGACCGTAACTTTCGACAGACCTTCTTCAATCTCTGCTGCGGCG	780		
DB	1035	ATCACATGCTGTGACCGACCGTAACTTTCGACAGACCTTCTTCAATCTCTGCTGCGGCG	976		
QY	781	GGTGACCGGATCTGTACCAACACATCTGTTGTTTTCGGGACCGCGGAGGTGTACATC	840		
DB	975	GGGACCGGCTGATGTTCCAGACCTGTTCTGGTCTTCGGGACCGCGGAGGTGTACATC	916		
QY	841	ATCATCTGCGCGCTTTCGCGATCATCAGCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTG	900		
DB	915	CTGATCTGCGCGCTTTCGCGATCATCAGCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTG	856		

gene
5167.5784
/gene="AGR_C_1406"
CDS
5167.5784
/gene="AGR_C_1406"
/note="CYTOCHROME C OXIDASE ASSEMBLY PROTEIN CTAG"
/codon_start=1
/transl_table=11
/product="AGR_C_1406p"
/protein_id="AAK86579.1"
/db_xref="GI:15155745"
/translation="MANMEARARQSRVSRSLVLCVFPFCAMIGMAYAAVPLYS
LFCRVYNGTTRVQYQSDVILDKINTVTPDANSNGLNMDRFDVKVKKVPGVFI
KVFKAIRGSPVATTAFTVFNTPMEAGYFNNKVECFETETTLQFGETLEMPVFFV
DPIAARETKSVHTLTSYTFYPAKTEKVPASLPVKTESGSKL"
5846.6721
/gene="AGR_C_1407"
CDS
5846.6721
/gene="AGR_C_1407"
/note="(AJ242592) cytochrome c oxidase subunit III"
/codon_start=1
/transl_table=11
/product="AGR_C_1407p"
/protein_id="AAK86590.1"
/db_xref="GI:15155746"
/translation="MADTHQKHNDHYIIDPSPWPLIASIGAFITFGVCVYMYLSGG
SPKLGALANPMLFVIGLVIVLYVYAWADTIKEANGSHTRVVSILHLYGMIMFI
ASVMPFVAFVAFDASLPFHEAIOASRLVGTGOWPKGLEVIDPWLPLXNTVIL
LLSGTCVTRAHALLHNDKGLISGLALTVALGVGFSTVQVYEHIAHPDPKNSIYGA
TFMAGTGFHFVGVGTVFLVCLFRAIAGGFTPKQHFGEAAWIHFVDVWDLFLF
FAYIMGGWGAPLHG"
6941.7369
/gene="AGR_C_1409"
CDS
6941.7369
/gene="AGR_C_1409"
/note="(AJ242592) hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_1409p"
/protein_id="AAK86581.1"
/db_xref="GI:15155747"
/translation="MYSRQPGCPREVRVLSNPNFAPVDPIKVGLKGCPCPCGNGR
LFDGFTPKPACAGCLDGDAGDGPVAVYMLIVGLVWGMALVQDPFAPPVWAH
VMLWLPVTVLSVLRLKLGIMIALQYRNNASEGRLDRE"
7338.8114
/gene="AGR_C_1411"
CDS
7338.8114
/gene="AGR_C_1411"
/note="(AF182952) Surfeit 1"
/codon_start=1
/transl_table=11
/product="AGR_C_1411p"
/protein_id="AAK86582.1"
/db_xref="GI:15155748"
/translation="MQAKGLIYNDVSHMKAAQRRVWFAPIVLALAILLGLTWQV
KRLYWKALMADIEERNASPAIDIEALIVKSGGEIYRKYRLSGTGDHTRERHFA
THOQTGYITLTLDAGRIILFVNRGFPVFEMKEAKRPDQGVSGEVTITGLARAPL
VAKPSSLLPNDIAKNIIFYWKDLAAMASSAEVPPERLVKLFVDNADNAPNGWPQGGV
TLIDLNNHLYQALTWYGLAALVIIVAGFAYFENGKAQGE"
8173.9219
/gene="AGR_C_1414"
CDS
8173.9219
/gene="AGR_C_1414"
/note="(AB006203) LytB"
/codon_start=1
/transl_table=11
/product="AGR_C_1414p"
/protein_id="AAK86583.1"
/db_xref="GI:15155749"
/translation="WNIAVSKPELTIRLCPGRFCAGVDRAIQIIVVLAIKAYGAPVYV
RHEIVHNYVVEGLKQALFVELHEI PAEHRQVPSVAFSGVPSKVPEDAQRNLF
YLDATCLPLVSKVQKQRNHRQGRHVLIHAGHPVEVITGMQLPEGVTSIVETVEDA
GVYEPVDRENLFVTTTSLSDDTAGVIAERLQERFPFPAIQAPAADSTCYATTNRQDAVK

gene
9312.10280
/gene="AGR_C_1416"
CDS
9312.10280
/gene="AGR_C_1416"
/note="HOMOSERINE KINASE (HK)"
/codon_start=1
/transl_table=11
/product="AGR_C_1416p"
/protein_id="AAK86584.1"
/db_xref="GI:15155750"
/translation="NAVYIDITDELRNPLTOYVGSLSYKGAEGVENSFNLLHTT
KPELILITVEKVENXNDJFPLGMQHLAAKSCPLPLPRKDGELJUGELSGRPAALI
SLEGMWLEKPEAKREVRVFKALAAHMLASEGFEIKRNPALSDVQWKLWMDKBERAD
EVEKGLBERIRPEIDYLAHWPKDLFAGVIHADFDQDNVFLFGBELSLIDIFYACND
LLAYDVISI CINAWCEKDGAVNTKGLLEGYSQVPLSEAELEALPLLSRGSALRF
FTRLYDLWLTTPAGALVVKDPLEYLRKLRPHRTIANVAEYGLAGE"
10277.10717
/gene="AGR_C_1417"
CDS
10277.10717
/gene="AGR_C_1417"
/note="Ribbonuclease H (E.C.3.1.26.4) Mutant With His 62
Replaced By Pro (H62p)"
/codon_start=1
/transl_table=11
/product="AGR_C_1417p"
/protein_id="AAK86585.1"
/db_xref="GI:15155751"
/translation="MKHVDIFDYGACSNPFGMGAVLRIGETEKELSGSEADTTNN
RMELAAISALNALXSPCEVDLYTDSAYKDGITKWIIFGKKKGKWKADNRPVKNVEL
WQALEAAQERHKVTLHFWKGHAGHPENERADELARKMEPFRR"

ORIGIN

Query Match 44.2% Score 740.4 DB 1 Length 10835;
Best Local Similarity 68.0% Pred. No. 1.5e-143;
Matches 1109; Conservative 0; Mismatches 511; Indels 12; Gaps 5;
QY 13 GCATTACGGCCCATACACCATATGAGCAAGGCTTTCACGCGTGTTCATGCG 72
Db 2025 GACGATCATTCGATCATTCATCCCATAGACCGGGCTTTTCGCCCGCTTCTGTCC 2084
QY 73 ACCAACACAGAACATCGTCTGCTATACCTTTAGCGGCTGTGTGTGCTTTCATT 132
Db 2085 ACCAACACAGAGCATCGGACGCTGTACCTGATCTTCGCCATCATGCGGCTATCATC 2144
QY 133 TCCGTCCTGTACCTCTACATGCGCCTTGAGCTGATGGATCCGGGTGTTCAGTACATG 192
Db 2145 GCGCGGGGCTTTCGGTCTGCTATGCGTATGAGCTGCAGAGCCGCGCATCCAGATCTTC 2204
QY 193 TGCCTTGAAGGCGACGCTGATCG---CGCATGCTCGCAGACATGTACGGCGAACGG 248
Db 2205 CACGGTCTGGCTTCCATGGTCTACGGCTTCGAGGGGATGCCGCCATCGACGCGGTAA 2264
QY 249 ACACCTGTGGAACGTCATGGTTACCTACCATGATTTCTGATGATGTTTCTTGTGGGTAT 308
Db 2265 -CATGTTTCAAGTGTTCACGACAGCGACGCGCTGATCATGATCTTCTTTCATGGTCAT 2323
QY 309 CCGCGCATTTCTCGTGTGTTTGGTAACTATCTGATGCGCTGCAATCGGGCTCCCGA 368
Db 2324 GCGCGCATGATCGCGGTTTTGCCCACTGGATGATCCGATCATGATCGCGCGCGCTGA 2383
QY 369 TATGGCCTTCCCGGTATGAACAACCTGCTGTTCTGGTGTTCATTTGCGGTGATCGCGAT 428
Db 2384 CATGGCTTTCGCGGCTT-AACTATTTCTTCTGGCTCATCGTTCGCGCTTCATCCT 2442
QY 429 GGGCGTGGCTTCGCTGTTCGACCGGGGTGACGCTGAGCTGGTGGTTCGGGGTGGTTG 488
Db 2443 GCTGCTCCTGCTCTTGTGCGAGGTCGGGGGTGTTATGCGGTGGTGGCGCTG 2502
QY 489 GGTTCGTGATCCCGCGCTGTCGACC---CGCAAGCTGCGCTATTCATGAGCTCGCGAT 545

Db 2503 GACCATGATCCGCTCTGTGACACAGTGGCATGCCCGGCCGCGGTGGATCTTGGCAT 2562
Qy 546 TTTCCGGGTTCACTGTGCGGTGCGCTCTCGATCATGGGCGCGCATCAACATGATCAGCAC 605
Db 2563 CTTCTCGTGCAGTGCAGCGTGCATCTCGATCCTCGCGGCCATCACTTCAACCCAC 2622
Qy 606 CTTCTTGAACATGCGCGGCCCGGCATCAGCTGCACAAAGTGCCTGTTCTTCGTGGTC 665
Db 2623 CATCTGAACATGCGCGGCCCGGCATCAGCTGCACAAAGTGCCTGTTCTTCGTGGTC 2682
Qy 666 GATCTTTATACAGGCTTGGCTGATCTCTCTGCGCGCTGCAGGTTCTGCTGGTGAATFAC 725
Db 2683 GGTCTGTGTCACCGGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2742
Qy 726 CATGCTGTGACCGACCGTAACTTCGGCAGCAGCTTCTTCAATCTCTCTGCGCGCGGTGA 785
Db 2743 CATGCTTCTGACCGACCGTAACTTCGGCAGCAGCTTCTTCTCTCTCTCTCTCTCT 2802
Qy 786 CCGGATCTCTACCAACATCTCTGTGTGTTCTTTGGGACCCGGAAGTGTACATCATCAT 845
Db 2803 TCCGATCTCTTACCAAGCATCTGTGTGTGTTCTTCTGCTTCTCTGCTTCTCTGCT 2862
Qy 846 TCTGCGCGGCTTGGCATCATCAGCCATGTCTGTGCACTTCTCGAAAGCGGTCTT 905
Db 2863 CTTGCGCGGTTTGGCATCATCAGCCATGTCTGTGCACTTCTCTCAAGACCGGTCTT 2922
Qy 906 CGGTTACCTGCGGATGTCTATGCAATGGTGGCAATCGGTCTTCTGGGTTCTGCTGCTG 965
Db 2923 CGGTTATCTCGGCATGTGTGCGCATGTGCGCATGTGCGGTCTGCGGTCTGCTGCTG 2982
Qy 966 GCGCACACATGTATACCGTGTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1025
Db 2983 GCGCACACATGTATACCGTGTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3042
Qy 1026 CACCATGTGTGATGCGGTGCGGACCGGATTAAGATCTTCTGTGATGCGGACGATGTG 1085
Db 3043 CACCATGTGTGATGCGGTGCGGACCGGATTAAGATCTTCTGTGATGCGGACGATGTG 3102
Qy 1086 GCGCGGCTGCTGTGAGTTCAAAATCGCGATGCTCTGCGGCTTGTGCTTATGCTGCTGT 1145
Db 3103 GCGCGGCTGCTGTGAGTTCAAAATCGCGATGCTCTGCGGCTTGTGCTTATGCTGCTGT 3162
Qy 1146 CACGTTGGGTGTGACCGGTATGCTGTGCGGCGGATGCTGCGGCGGCTGCGGCGGATCA 1205
Db 3163 CACGTTGGGTGTGACCGGTATGCTGTGCGGCGGATGCTGCGGCGGCTGCGGCGGATCA 3222
Qy 1206 CGACACCTATTACGTGTGCGGCTTCTGATGCTGCTGCGGCGGATGCTGCTGCTGCTGCT 1265
Db 3223 CGACACCTATTACGTGTGCGGCTTCTGATGCTGCTGCGGCGGATGCTGCTGCTGCTGCT 3282
Qy 1266 GATCTTCCGCGGTATCTACTTTTACATGCGGAGTCTTCCGCGCGGCTTCCCGGAAATG 1325
Db 3283 GATCTTCCGCGGTATCTACTTTTACATGCGGAGTCTTCCGCGCGGCTTCCCGGAAATG 3342
Qy 1326 GCGTCAAAAGTGTGACTTCTGCACTTCTTCCGCGGAGTCTTCCGCGCGGCTTCCCGGAA 1385
Db 3343 CATCGGCAAGTGTGACTTCTGCACTTCTTCCGCGGAGTCTTCCGCGCGGCTTCCCGGAA 3402
Qy 1386 GCATCTTCTGGGAGTGTGCGGCGGCTTACATGCACTATCCCGGAGCGCTTCCGCGGAA 1445
Db 3403 GCATCTTCTGGGAGTGTGCGGCGGCTTACATGCACTATCCCGGAGCGCTTCCCGGAG 3462
Qy 1446 GCTGTGGAACAAAGTGTGCGGCGGCTTACATGCACTATCCCGGAGCGCTTCCCGGAG 1505
Db 3463 CGGCTTGGACATGTATGCTCTACGCTTCTATAT---CTCGCGGCTTCCCGGAGCGAT 3519
Qy 1506 CATGCTGATCTTGTGCTATACGCTGTGTGCGGCGGCGGAGACCGGCTTCCCGGAGCG 1565
Db 3520 CTTCTCTTCTGCGGCTTCTGCGGAGCGCTTCTGCGGAGCGGATCGCGGCTTACCAATCC 3579
Qy 1566 GCGCGAATTCGCGGATACGCTGGAATGAGCGCTTCCGCGGCGGCTTCCCGGAGCGGAT 1625
Db 3580 GCGTGAAGGTGCGGAGCGCTTGAATGAGCGCTTCTTCTGCGGCGGCTTACCAACAGTG 3639

QY 1626 CGAAACGCTGCC 1637
Db 3640 GGAACGCTTCC 3651

RESULT 10
AB014356

LOCUS
DEFINITION

Brucella suis 1330 chromosome I section 42 of 190 of the complete

genome.

ACCESSION

AE014356 AE014291

VERSION

AE014356.1 GI:23347257

KEYWORDS

SOURCE

Brucella suis 1330

Brucella suis 1330

Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

Brucellaceae; Brucella.

1 (bases 1 to 10154)

Paulsen, I., Seshadri, R., Nelson, K.E., Eisen, J.A.,

Heidelberg, J.F., Read, T.D., Dodson, R.J., Umayam, L.M.,

Beanan, M.J., Daugherty, S.C., Deboy, R.T., Durkin, A.S., Kolonay, J.F.,

Madupu, R., Nelson, W.C., Ayodeji, B., Kraul, M., Shetty, J., Malek, J.,

Van Aken, S.E., Riedmuller, S., Tettelin, H., Gill, S.R., White, O.,

Salzberg, S.L., Hoover, D.L., Lindler, L.E., Halling, S.M., Boyle, S.M.,

and Fraser, C.M.

The Brucella suis genome reveals fundamental similarities between

animal and plant pathogens and symbionts

Proc. Natl. Acad. Sci. U.S.A. 99 (20), 13148-13153 (2002)

22247741

12271122

2 (bases 1 to 10154)

Paulsen, I., Seshadri, R., Nelson, K.E., Eisen, J.A., Heidelberg, J.F.,

Read, T.D., Dodson, R.J., Umayam, L.A., Brinkac, L.M., Beanan, M.J.,

Daugherty, S.C., Deboy, R.T., Durkin, A.S., Kolonay, J.F., Madupu, R.,

Nelson, W.C., Ayodeji, B., Kraul, M., Shetty, J., Malek, J.A., Van

Aken, S.E., Riedmuller, S., Tettelin, H., Gill, S.R., White, O.,

Salzberg, S.L., Hoover, D.L., Lindler, L., Halling, S.M., Boyle, S.M.,

and Fraser, C.M.

Direct Submission

Submitted (14-AUG-2002) The Institute for Genomic Research, 9712

Medical Center Dr., Rockville, MD 20850, USA

Location/Qualifiers

1. 10154

/organism="Brucella suis 1330"

/mol_type="genomic DNA"

/strain="1330"

/db_xref="taxon:204722"

/chromosome="I"

135..1034

/locus_tag="BR0463"

135..1034

/locus_tag="BR0463"

/note="identified by similarity to OMNI:NTL02ML6198; match

to protein family HMM PF03755; match to protein family HMM

TIGR00255"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein TIGR00255"

/protein_id="AA029406.1"

/db_xref="GI:23347258"

/translation="VALQSTGTFARHAQGMADGEARIVWEVRSVNGKGLDLRLRLP

QGLSEVHPVRSLLARHFRGNFQASLTVERSETQGFAPINQPMALVFLKGLAEQLEOK

YGLAPASVDGIIALRGIIIDQSRADEBEERALEATIIVAAPEGALKIAEARRQGEKA

LFTILSHLTJERLTADARNPSPRATKAIKARLAGOVALLDATREIDETRLYCEAA

FLATKADIQBEDRLSTHVASARKLLGEGPVGRKLDLFLSQEFNREANTLCSKSNAS

ITAIGLKRAVDQDFREQVQNL"

1043..1705

/gene="gmk"

/locus_tag="BR0464"

1043..1705

/gene="gmk"

/locus_tag="BR0464"

gene

CDS

gene

CDS

Qy	5	CAGACGCCGCATFTACGGCCATGACCAACATCAGAAAGCAAGCTTCTTTCACGGCGTGGT	64
Db	5348	CAGCAGTCAACGACATGTCGCCACGACACCAACGACGCGATGGCTGGGTTTGTTGGG	5407
Qy	65	TCATGTGACCAACACAAAGACATCGTCTGCTATACCTTGTAGCGGCTGGTGTCTTG	124

Db 5408 TATACTCGACCAATCATAAAGACATCGGTACCTGTACTGATTTTGGCAATCATCGCG 5467
 Qy 125 GTTTCATTTTCGTCCTGTTCAACCGCTTACATAGCGCTTGAGCTGATGATCGCGGTTC 184
 Db 5468 GCATCATCGCGGGCGGCTTTTCCATCGCATCGGTGCGGAACCTCAGGAGCGGGCATCC 5527
 Qy 185 AG---TACATGTGCTTGAAGCGCAGCTCTGATCGCGGATGCTCGACACATGACGG 241
 Db 5528 AGATTTTCCATGTCGCGCAGATGGTTTACGGTGTGAAGCGATGCGCCCTTGAATG 5587
 Qy 242 CGAAGCGACACCTCTGGAACGTCATGTTTACCTACCATGATTTCTGATGATGTTCTTTG 301
 Db 5588 CCGCAAGACATGTTCAACGTGTTACCTCGCGGCACGCGCTCGTCATGATCTTCTTCA 5647
 Qy 302 TGGGTATCCCGCATGTTGCGTGGTTTGGTAACTATCTGATGCGCGCTCAATCGCGG 361
 Db 5648 TGGTCATGCTGCGCTGATCGCGGTTTCGCAACTGGATGGTGGCGCTCATGATCGCGG 5707
 Qy 362 CTCGGATATGCGCTTCCCGGTATGAACAACTGTCGTTCTGGCTGTTCAATGCGGTA 421
 Db 5708 CGCGGACATGGCGTTCACGCGATGAACAATTTCTTCTGGCTGCTGCGCGCGCGC 5767
 Qy 422 CCGGATGGGGTGGCTTGGTGTTCGACCGGGCGGTGACGCTACGCTGGGTTCGGGG 481
 Db 5768 TGACGCTGCTGCTGATCTCGCTGTTTCATGCGCGTCCGCGAGGTGGCTGGGGCGCGG 5827
 Qy 482 TTGGTTGGGTCTCTACCGCGCGCTGCGA---CCGCGGAAGCTGGGTATTTTCATGAGAC 538
 Db 5828 CGGCTGAGCGCTATCTCGCGCTTCTGACCTCCGCGCAGCGCGCGCGCGGTGAT 5887
 Qy 539 TCGGATTTTCGGGTTCACTTGTGCGGTGCTCTCTGATCATGGCGCGGATCAACATGA 598
 Db 5888 TCGGATTTTCGGCATCCACATTTCCGGTGCCTGCTGATCTCTCGCGCGGATCAACTTCA 5947
 Qy 599 TCAGACCTTCTTGAACATCGCGCGCGCGCGCATGAGCTGCAACAGTCCGCTTCTTCT 558
 Db 5948 TTACAGCATCTGAAATGCGCGCGCGCGCATGAGCTGCAACAGTCCGCTTCTTCT 6007
 Qy 659 CGTGGTCGATTTTATCACGGCTTGGCTGATCTGCTGGCGTCCGCGTTCGCTGGGTG 718
 Db 6008 CCTGGCGCTTCTGGGTACGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6067
 Qy 719 CAATCACATGCTGCTACGACGCGTAATCTTCGCGACGCTTCTCAATCTCTGCTGGG 778
 Db 6068 GTATCACATGCTGCTTACGCGCGCAATTTCCGCGACGCTTCTTTCGCGCTGATGG 6127
 Qy 779 GCGGTGACCGGATCTCTACCAACACATCTGCTGGTCTTTTGGCACCGCGAAGTGACA 838
 Db 6128 GCGGCGACCGATCTCTACGACGCTGTTCTGGTCTTCGCGCATCGGAGTCTATA 5187
 Qy 839 TCATCATCTGCGCGCTTGGCATCATAGCATGCTGCTGCACTTCTCGCAAAAGC 898
 Db 6188 TCCTGATCTCTCGCGGCTTCCGATGCTCAGCATGCTGCTCCTGCTGCTGCTGCTG 6247
 Qy 899 CGGTCTTCGTTTACCTCGCATGCTTATGCAATGTTGCAATCGGTGTTCTGGGTTTG 958
 Db 6248 CGATCTTCGTTTATCTCGCATGCTTACCGCATGTTGCAATCGGTGCTGCTGGGTTTG 6307
 Qy 959 TCGTCTGGCGCACCAATGATACACGCTTGGTATGCTGCTGACCCAGCAATCTACTTCA 1018
 Db 6308 TCGTCTGGCGCACCAATGATACACGCTTGGTATGCTGCTGACCCAGCAATCTACTTCA 6367
 Qy 1019 TCGTGGCACCATGTTGATCGCGTCCGCGCATTTAAGATCTTCTGTTGATCGCA 1078
 Db 6368 TCTTCGCGCATGTTGATCGCGTCCGCGCATCAAGATCTTCTGTTGATCGCA 6427
 Qy 1079 CGATGTTGGCGCGCTCGGTGAGTTCAATTCGCGCATGCTCTGGGCTTTTGGCTTTATGT 1138
 Db 6428 CGATGTTGGGTTGCTCGATTTCTTCCGCGCGCGCATGCTCTGGGCTTCTGCTTCACT 6487
 Qy 1139 TCGTGTTCACGTTGGTGTACCGGTATCGTGTGCGCGCAAGCGGTCTGGAACGCTG 1198
 Db 6488 TCTGTTTACCATCGCGGTTGACGGGTGTGAGCTTGCAGCTTGCAGCTTGCAGCGCTG 6547

Qy 1199 CATATCAGACACCTATTACGTGGTGGCGCACTTCCATTATGTATGTGCTGGGTGCGA 1258
 Db 6548 CATTGACGATACCTATTATGTGGTGGCGCACTTCCACTATGTGCTGCTGGCGCTG 6607
 Qy 1259 TCTTTGGCATCTTCGCGGTATCTACTTTTACATGCGCAAGTTCTCGGCGCGCTTCC 1318
 Db 6608 TCTTTCGCACTTTTCGCGGTGTTACTACTGTTTCCGGAAGATGAGCGGTACATGATA 6667
 Qy 1319 CGAATGCGCTCAAAAGCTGCACTTCTGCACTTCTTTCATCGGTGCGAACCTCAGTTCT 1378
 Db 6668 ACGAGTTCACGCAAAAGCTGCACTTCTGCTGGTCACTGCTGCTGGTGAACCTGCTTCT 6727
 Qy 1379 TCCCGCAGCACTTCTGCGAGTCAAGGTATGCGCGCGCTTACATGCACTATCCGAG 1438
 Db 6728 TCCCGCAGCACTTCTGCGGTCTGCGCGCATGCGCGCGCTTACATGCACTATCCGATG 6787
 Qy 1439 CTTTTCGCGCTGTGGAACAAAGTCTGCTCTATGGTGGTCTTCTGCGCTTCTGCTTCC 1498
 Db 6788 CTTATGCGCGCTGGAACATGATTTGCTCATATGGCTCTCTACATTTTCAAGCTTTGCCGTG 6847
 Qy 1499 TCTTCTTCATGCTGATCTTTGCTATACGCTGCTGCTGGCGCGCGGAGACCGTCCGA 1558
 Db 6848 TGATCTTCTCTATAAC---GTCTTCGAGGCGTTCGCGAAGAGCGTGAAGCGGTGCAA 6904
 Qy 1559 ACCCGTGGGCGAATTCGCGATACGCTGGAATGAGCGCTGCCATCACCGCTCCGCGCC 1618
 Db 6905 ATCCGTGGGCGATGCGCGACCGCTGGAATGCGAGCTTCTTTCGCGCGCGCTTCC 6964
 Qy 1619 ACACGTTCCAAACGCTGCC 1637
 Db 6965 ACCAGTGGGAACAGCTGCC 6983
 RESULT 11
 AE009583/c
 LOCUS
 DEFINITION
 Brucella melitensis 16M chromosome I, section 140 of 195 of the complete sequence.
 ACCESSION
 AE009583 AE008917
 VERSION
 AE009583.1 GI:17983463
 KEYWORDS
 SOURCE
 ORGANISM
 Brucella melitensis 16M
 Brucella melitensis 16M
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Brucellaceae; Brucella.
 1 (bases 1 to 11404)
 DelVecchio, V.G., Kapatral, V., Redkar, R.J., Patra, G., Mujer, C., Los, T., Ivanova, N., Anderson, I., Bhattacharyya, A., Lykidis, A., Reznik, G., Jablonski, L., Larsen, N., D'Souza, M., Bernal, A., Mazur, M., Goltzman, E., Selkov, E., Elzer, P.H., Hagijs, S., O'Callaghan, D., Letesson, J.-J., Haselkorn, R., Kyripides, N. and Overbeek, R.
 The genome sequence of the facultative intracellular pathogen Brucella melitensis
 Proc. Natl. Acad. Sci. U.S.A. 99 (1), 443-448 (2002)
 11756688
 2 (bases 1 to 11404)
 DelVecchio, V.G., Redkar, R.J., Patra, G. and Mujer, C.
 Submitted (13-NOV-2001) Institute of Molecular Biology and Direct Submission
 Medicine, University of Scranton, Scranton, PA 18510, USA
 3 (bases 1 to 11404)
 Elzer, P.H. and Hagijs, S.
 Submitted (13-NOV-2001) Department of Veterinary Science, LSU Ag Center, 111 Dairymple Building, Baton Rouge, LA 70803, USA
 4 (bases 1 to 11404)
 Kapatral, V., Los, T., Ivanova, N., Anderson, I., Bhattacharyya, A., Lykidis, A., Reznik, G., Jablonski, L., Larsen, N., D'Souza, M., Bernal, A., Mazur, M., Goltzman, E., Selkov, E., Haselkorn, R., Kyripides, N. and Overbeek, R.
 Direct Submission
 TITLE
 JOURNAL
 PUBLISHED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE

JOURNAL Submitted (13-NOV-2001) Integrated Genomics, Inc., 2201 W. Campbell
Park Drive, IL 60612, USA
REFERENCE 5 (bases 1 to 11404)
AUTHORS Letesson J.-J.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2001) Unite de Recherche en Biologie Moleculaire,
Laboratoire d'Immunologie et de Microbiologie, Universite of Namur,
61 rue de Bruxelles, Namur 5000, Belgium
REFERENCE 6 (bases 1 to 11404)
AUTHORS O'Callaghan, D.
TITLE Direct Submission
JOURNAL Submitted (13-NOV-2001) Faculte de Medecine, INSERM U431, Avenue
Kennedy, Nimes 30900, France
FEATURES
source
1..11404
/organism="Brucella melitensis 16M"
/mol_type="genomic DNA"
/strain="16M"
/db_xref="taxon:224914"
/chromosome="1"
complement(85..834)
/gene="BMEI1460"
complement(85..834)
/gene="BMEI1460"
/codon_start=1
/transl_table=11
/product="SURPEIT LOCUS PROTEIN 1"
/protein_id="AAL52641.1"
/db_xref="GI:17983464"
/translation="MTLQTSRRFPWGLVLSGIALVLLSGTWQVRLMKKALLA
STEQRHPPPLAEEMKIQEGTVEYRPTVSGTFHQGERFLATYKGEAGFYV
TPLEDRGRVLRGFEYKQDPATPAGELAAAGPVTGLARDPLSKPSPFLVD
NDIAKNIYKWDMAAMASAGLPDLGRVVPFVDADSTPNPGLPIGGVTIIDPNNH
LQAYTWGLALGIVGTWLRVYKSGRG"
complement(831..1289)
/gene="BMEI1461"
complement(831..1289)
/gene="BMEI1461"
/codon_start=1
/transl_table=11
/product="2INC-FINGER PROTEIN"
/protein_id="AAL52642.1"
/db_xref="GI:17983465"
/translation="MAAFCSQVAFQSRNCRCFRGRKNAMPEHGHYPQVDPIRSGIAG
RCRCGEGKLFNGFLTAPRCTAGLDYGFADSGDPAFVILIGFVVGWALMEV
NIGPLVWHFILWPLAILSLVAMRWIKILINQHRKAEQCRLDNSK"
complement(1328..2272)
/gene="BMEI1462"
complement(1328..2272)
/gene="BMEI1462"
/EC_number="1.9.3.1"
/codon_start=1
/transl_table=11
/product="CYTOCHROME C OXIDASE ASSEMBLY PROTEIN CTAG"
/protein_id="AAL52643.1"
/db_xref="GI:17983466"
/translation="MRFTLQAGSAREEPRLNGRTGMDADVHOKNDYHIIAPSPFPL
SSVGFVLAIGIAMRWYINAGELPFFGLNVVTPWIFGLAIVLYCMFGWSDTIKE
GHEGHTRVSLHLRYGMIMFIASEVMFFVAFWAFDASLPFNHGVQVRAEFTGGV
WPKGTGLDPLHLPLNTVILLSGTLTWAHALLHNDKRGMITGLAITVALGLVF
SCVQAEVYTHAPFKDSTYGATFFMATGFHGFHVLIGTIFLLVCLFRAIGDFTPKQ
HGFGEAAAYWHFVDVWVLFLEAFIYVWGCGAPMHEG"
complement(2415..3128)
/gene="BMEI1463"
complement(2415..3128)
/gene="BMEI1463"
/codon_start=1
/transl_table=11
/product="CYTOCHROME C OXIDASE ASSEMBLY PROTEIN CTAG"
/protein_id="AAL52644.1"
/db_xref="GI:17983467"
/translation="MPSRRRFSFCWFTSARWPRWAPACSGRSDNRBERMTDQGENE
KKQRNSNATIAVCLSFVFCMIGAAVASVELYRIFQCVTGYGTTQRVEQYSDTILDK

TIKVRFDANIANGLPWDFKPMOREVTVRIGETTMIKYEAHNLFGSETYGRASFNVAPG
RAGAVNKVCEFCFTDNTLKGEDLELPWFVFDVPEFVNDPLDKVKITLTSYTFPI
DAPKVVAKAVGSTRNGG"
complement(3225..4172)
/gene="BMEI1464"
complement(3225..4172)
/gene="BMEI1464"
/EC_number="2.5.1.1"
/codon_start=1
/transl_table=11
/product="PROTOHEME IX FARNESYLTRANSFERASE"
/protein_id="AAL52645.1"
/db_xref="GI:17983468"
/translation="MSLVKNTASBEDAFALSEATARDYLVLKPRVMSLVVFTGLVGL
VLAPGHMNPDLAVISILCIAVAGASGALNMWYDADIDAVMKRTRKRPIAGIIAPNQ
VLARGLTISAFSVMTLGLMVNMLAALIAFTIFFVAVIYTMWKSTPQNIIVIGGAAG
APPMITGMAATGEITWDSLVLFPMIFLWPPHVALSFTTNDYEAAIIPWNPVKG
ELSTRQALFYAVLAPVGVLPWVMPGPMGYGVUSTLGLAFVYVAMWLWAADSQPK
MLAAARKLFRFSLYLAGIFAVILLFEALTFLKLLAAFGVF"
complement(4400..6058)
/gene="BMEI1465"
complement(4400..6058)
/gene="BMEI1465"
/EC_number="1.9.3.1"
/codon_start=1
/transl_table=11
/product="CYTOCHROME C OXIDASE POLYPEPTIDE I"
/protein_id="AAL52646.1"
/db_xref="GI:17983469"
/translation="MAGTAARHGHADHDKPHGVWVWYSTNHKDGITLYLIPAILAG
TIGGALSIAARAELOPQGIQIHLGAQVYGVGGDAALDAGKHMENVFTSAHALVMIF
FWMPALIGGFANWYPLMIGAPDMAFPRMNNISFWLLPALTLILLISLFMPGPGAGW
GGGGWTLVPPFTSGQGPAPVDFAILA.HISGASSILGAINPITILLNKRAGMTLH
KMFLEWALVTLFALLLSLPLVLAGITLWLLDNFTGTTFFAPDGGDPLIYOHLEWF
FGHEVITLQPGFIVSHVSVTSFRKPIFGYLGMAVAVAGVGVGVVVAHHMYTVG
LSLDTORYFVFAVIAVPTGIKIFSWIATMGGISFRAPLMAIGFTFLFTIGVT
GVLANAGLDRLDHTVYVVAHPVLSGAVFAI PAGWYMPKMGSGYMNFEFTAKL
HWVTFVGNLVFPQHFELGAMPREYIDYDPAVAGMNMISYSGSVISGFAVLIFLY
NVFEAKKREAGANPWGGATLEWQLSSPPPHQWELPRIK"
complement(6196..7158)
/gene="BMEI1466"
complement(6196..7158)
/gene="BMEI1466"
/EC_number="1.9.3.1"
/codon_start=1
/transl_table=11
/product="CYTOCHROME C OXIDASE POLYPEPTIDE II"
/protein_id="AAL52647.1"
/db_xref="GI:17983470"
/translation="MNADGKAVLKTGVQPERAKRRLVVDKIIVATRGAGGALAVL
WTTGAALADQPERPWEKRFQDAATGIAEQIHWFRYTLWFTIITLLVGLLVVWVLRF
RASVNPSTKSHNTALEVITVGVPIILLFLAVSPOLLTAQYSPEDPTLTVKATGY
QWNGYBYQDNPVSDSVILKDDDDRAAKGDEBRAPRLTVDNVEVWVPVGTVRL
VTASDVYHSWTIFAFGVNDVAVPGRINETWFKADKGLYTGQCSSELCKGKHGMPPIAV
RVVSKDQYQWLSAATDLFGANKALLQAESEGSANDKVAAGL"
7514..8110
/gene="BMEI1467"
7514..8110
/gene="BMEI1467"
/codon_start=1
/transl_table=11
/product="Hypothetical Protein"
/protein_id="AAL52648.1"
/db_xref="GI:17983471"
/translation="MIORAALAAAGLSAAATLPSAQNNTPTGQLPAPQPTPA
EPQTTTSATQCSGALXSQHGASILCDDTPAGAKTEOCALIONVAAKRPELGLSVV
LKTADNKARILRVLAPGLVILPGLNGLNVDGKDIGRAYFVRCFEDGCGYAEVILBDDLL
KTLNGKAATFVFTPEEGVGVPLDKLKGEGFDALP"
8251..9663
/gene="BMEI1468"
8251..9663
/gene="BMEI1468"
/codon_start=1

```

/transl_table=11
/product="TLDD PROTEIN"
/protein_id="AAL52649.1"
/db_xref="GI:17983472"
/translation="MKSLIDSPDTRGRDHIERIVAESLKSDDGELFLEVRERBALVFD
NGLKTSQFNQDQFGKRSVAGEAIGYAHAGDLSLAALKRASDAVSATLKGKGGTYSP
APATNHLITDENIPICAPGFDKAVQQLQIDAYLRADTKVQSVSLACSQQQVEI
LRADGHVPRDIRPMARLSVAVAGSDRQBSGYTLGGRKGFEFTTTSWOFHAADEA
LRQALVLEAI PAPAGTFDVLANGWPGVMLEHAGVLEGGDFNRKRTSAFAGLLQOQ
VAARGVTVDGRTIARERGLTDDGTPNRTVLIEDGKLFNMDODRONARLGMGEA
TNGRRRESYAHAPMPRTNMYLSGDKTPEELIASIRKGIYAVSFGSGVDLITSGKEV
FGCTEAYMIEDGKVGADPKAMLLINGPDMKRLISMIGNDKMLDTSNGCGGQGWVP
VGVGQHLRDQITVGGTAV"
complement(9690..10457)
/gene="BMEI1469"
complement(9690..10457)
/gene="BMEI1469"

```

```

Query Match      44.08; Score 736.6; DB 1; Length 11404;
Best Local Similarity 66.98; Pred. No. 9.1e-143;
Matches 1096; Conservative 0; Mismatches 534; Indels 9; Gaps 3;

Qy 5 CAGACGGCGCATTCACGGCGCATGACACCATGAGAGCAAGGCTTCCTCACGCGCTGGT 64
Db 6048 CAGCAGCTCAGGATGGTGCCACGACGACACAGCGCATGGCTGGGTTGG 5989

Qy 65 TCATGTCGACCAACCAAGACATGGTCTGTATACCTGTAGCGGTGGTGTGGT 124
Db 5988 TATACTCGACCAATCATAAAGACATGGTACCTGTACCTGATTTTGAATCATCGCG 5929

Qy 125 GTTTCATTTCCGTCCTGTTTCCACCGTCTACATGCGGCTTGAGCTGATGGATCGGGGTTC 184
Db 5928 GCACCATCGCGGCGCTTTTCATCGCATGGATGGTGCGGACCTCAGAGCGCGGATCC 5869

Qy 185 AG---TACATGTGCTTGAAGGCGACGCTGTGATCGCGATGCTCGCAGACATGACG 241
Db 5968 AGATTTTCCATGTCTGGCGCGATGTTTACCGTGTGAAAGCGATGCGGCCCTTGATG 5809

Qy 242 CGAAGCGACACCTGTGGAAGCTCATGGTTACCTACCATGGTATCTGATGATGTTCTTTG 301
Db 5808 CCGGCAAGCACATGTTCAAGCTGTTCACTCCGCGCACGCGCTGTCATGATCTTCTTCA 5749

Qy 302 TGGGTATPCCCGCATTTGTCGGTGGTGGTAACTATCTGATCGCGCTGCATAATCGGCG 361
Db 5748 TGGTCATGCTGCGCTGATCGGCGGTTTCCGCAACTGGATGGTGGCGCTCATGATCGGCG 5689

Qy 362 CTCGGATATGGCTTCCCGCGTATCAACACTGCTGCTTCTGGCTTCATGCGGTA 421
Db 5688 CGCGGACATGGGCTTCCACGATGAAACAATTTCTTCTGGCTGTCGCGCGCGCGG 5629

Qy 422 CCGCGATGGCGGTGGCTTCTGCTTTCGACCGGCGGTGACGCTCAGCTGGGTTGCGGCG 481
Db 5628 TGACGCTGCTGTGATCTCGCTGTTTCATCGCGTCCGCGAGTGGCTGGGCGCGCGCG 5569

Qy 482 TTGGTTGGTTCGTACCGCGCTGCGA---CCGCGAAGCTGGCTATTCATGAGACC 538
Db 5568 GCGGCTGGACGCTCTATCGCGCGTTTCGACCTTCGCGCGCGCGCGCGCGGCGGAT 5509

Qy 539 TCCGATTTTCCGGGTTCATTTGTCGGGTGCTCTCGATCATGGCGCGATCAACATGA 598
Db 5508 TCGCCATCTCGGCATCCCATTTTCGGTGGCTCTCGATCTCGCGCGATCACTTCA 5449

Qy 599 TACGACCTTTTGAACATGGCGCGCGCGCGGATGACGCTGCACAAAGTGGCTGTCT 658
Db 5448 TTACGACGATCTCTGAACATCGCGCGCGCGGATGACGCTGCACAAAGTGGCTGTCT 5389

Qy 659 CGTGGTGCATCTTTATCAGGCTTGGCTGATCTGCTGCGGCTGCGGCTTCTGGCTGTG 718
Db 5388 CTTGGGCGCTTCTGTTAGGCGTTCCTGCTGCTTCTGCTGCTTCTGCGGCTTCTGCGAGTG 5329

Qy 719 CAATACCATGCTGTGACCGACCGTAATCTCGGACGACCTTCTCAATCTGCTGGCG 778
Db 5328 GTATACCATGCTGTACCGACCGTAATCTCGGACGACCTTCTGCGGCTGTGCGG 5269

```

```

Qy 779 GGGGTGACCCCGATTCTGTACCAACACATCTCTGTGTTCTTTGGCACCCGGAAGTGTACA 838
Db 5268 GGGGCGACCCGATCTCTTACCAGACCTGTCTGTGTTCTTCGGCCATCCGGAAGTCTATA 5209

Qy 839 TCATCATTTCTCCCGGCTTTTGGCATCATACGCCATGTGCTGACCTTCTCGAAAAAGC 898
Db 5208 TCCTGATCCAGCGGCTTCGGCATCGTCAGCCATGTGCTCTCCACCTTCTCGCGCAAGC 5149

Qy 899 CGGCTTCGGTTACTGTCGGCATGCTCTATCAATGTTGGCAATCGGTGTTCTGGGCTTTG 958
Db 5148 CGATCTTCGGTTATCTCGGATGGCTTACGCCATGGTGGCAATCGGTGTCGTGGGCTTTG 5089

Qy 959 TCGTCTGGGCGCACACATGTACACCGGTTGATGTGCTGACCCAGCAATCTCTACTTCA 1018
Db 5088 TCGTCTGGGCGCACACATGTACACCGGTCGCGCTTTCGCTCGATACACAGCGCTACTTCG 5029

Qy 1019 TGCTGGCCACCATGGTGTATCGCGTGGCGACCGCATTAAGATCTTCTCTGTTGATCGCCA 1078
Db 5028 TCTTGGCCACCATGGTGTATCGCGTGGCGACCGCATTAAGATCTTCTCTGTTGATCGCCA 4969

Qy 1079 CGATGTGGGCGGCTCGGTTGAGTTCAAATCGCGATGCTCTGGGCTTTGGCTTTATGT 1138
Db 4968 CGATGTGGGCTGGCTCGATTTCTTCCGCGCGCTCATGCTCTGGGCTATCGGCTTCACT 4909

Qy 1139 TCCTGTTACCGTGGTGGTGTGACCGGTATCGGTGTCGCGCCCAAGCGGTCTGACCGGTG 1198
Db 4908 TCCTGTTACCATCGCGGTGTGACCGGTGTGCAAGTTGCAAAATGCGCGTCTCGACCGTG 4849

Qy 1199 CATATCAGCACACTATTACGTGTGGCGACTTCCATTTATGATGATGCTGCGTGGTSCGA 1258
Db 4848 CATTCGACGATACCTATTTATGTTGGCGCACTTCCACTATGCTGCTGCTTGGCGCTG 4789

Qy 1259 TCTTTGCGATCTTCGCGGTATCTATTTTACATGCGCAAGTTCTCGGGCGCGCTTTC 1318
Db 4788 TCTTTCGCGATCTTTCGCGGTGTTACTTACTGTTTCCGAAAGATGAGCGGTACTATGTATA 4729

Qy 1319 CGGAATGGGTGCAAGCTGCACTTCTGGACCTTCTTACGCGGTGCGAAGTCACTTCT 1378
Db 4728 ACAGATTACCGCAAGCTGCACTTCTGGGTGAGTTCTGCGGTGTAACCTGGTCTTCT 4669

Qy 1379 TCCCGCAGCACTTCTCGCGACGTCAGGATGTCGCGCGCTTACATCGACTATCCGGAAG 1438
Db 4668 TCCCGCAGCACTTCTCGCGCTCGCGGCAATGCGCGCGCTTACATCGACTATCCGGAAG 4609

Qy 1439 CTTTCGCGCTGTGAAACAAAGTCTGCTCTATGTTGGTGTCTTGGGCTTGGCTCGTTC 1498
Db 4608 CTTATGCGGCTGGAACATGATTTGCTCATATGGCTCCTACATTTTTCAGGCTTTGCGGCTG 4549

Qy 1499 TGTTCCTTCATCGTGTATCTTGTGTATACGCTGGTGTGTCGCGCGCGAGACCGCTCGGA 1558
Db 4548 TGAATCTTCTCTATAAC---GTCTTCGAGCGGTTCGCGAAGAGCGTGAAGCGGTGCAA 4492

Qy 1559 ACCGTTGGGCGAATTCGCGATACGCTGGAATGAGCGTGCATCACCGCTTCCGCGCC 1618
Db 4491 ATCCGTTGGGCGATGCGCGACCAACGCTGGAATGGAGCTTTCTTCGCGCGCGCGCTTTC 4432

Qy 1619 ACAGTTTCGAAACGCTGCC 1637
Db 4431 ACCAGTGGGAACAGCTCCC 4413

```

```

RESULT 12
AP003011
LOCUS
DEFINITION Mesorhizobium loti DNA, complete genome, section 18/21.
ACCESSION AP003011 BA000012
VERSION AP003011.2 GI:114026664
KEYWORDS
SOURCE Mesorhizobium loti
ORGANISM Mesorhizobium loti
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Phyllobacteriaceae; Mesorhizobium.

```

REFERENCE 1 (sites) Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S., Watanabe,A., Idesawa,K., Ishikawa,A., Kawashima,K., Kimura,T., Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A., Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M., Takeuchi,C., Yamada,M. and Tabata,S.
Complete genome structure of the nitrogen-fixing symbiotic bacterium Mesorhizobium loti
DNA Res. 7 (6), 331-338 (2000)
21082930
MEDLINE
PUBMED 11214968
REFERENCE 2 (bases 1 to 346510)
Kaneko,T.
Direct Submisison
Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research, Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:kaneko@kazusa.or.jp, /rhizobase/,
URL:http://www.kazusa.or.jp/rhizobase/,
Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
On May 11, 2001 this sequence version replaced gi:11994986.
COMMENT Location/Qualifiers
FEATURES
source
1..346510
/organism="Mesorhizobium loti"
/mol_type="genomic DNA"
/strain="MAFF303099"
/db_xref="taxon:381"
complement(242..1264)
/gene="ml17087"
complement(242..1264)
/note="similar to lipopolysaccharide core biosynthesis glycosyl transferase lpse"
codon_start=1
/transl_table=11
/protein_id="BAB53261.1"
/db_xref="GI:14026665"
/translation="MAGALACAVRLVSDVHLPRILAIADRPCWAIDRKTQMLRVLER RFEIVQRQHDVIESDVAADIVLIFYWFOIAKULPESVLARSDDLVIIGCSHREL EGEFREPGLATRLHRAVFNRRKLEHYAPLLRIPVHTYTPNGVDYTFCKRPEPPQ YGSELRVWAGSLGNHGAHGFHDVIEPAVAAPGARILLTAIREQHWNRHDEMDFY RDLVVVCASRSEGTNPCLAAACGIPLVTVRCVMPPELVQDGDNGLFFDGTAEGLA DXLALRTPSRLSRMAARMETIRGMDWRVQAEYAHMFTSLITAGGAVPEVRAAS DVFQQT"
complement(1228..2265)
/gene="ml17088"
complement(1228..2265)
/gene="ml17088"
/note="similar to lipopolysaccharide core biosynthesis glycosyl transferase lpse"
codon_start=1
/transl_table=11
/protein_id="BAB53262.1"
/db_xref="GI:14026666"
/translation="MALIAGEHLQGGVSTKCGALAPPSLIRLVI SAKFSNCAILV ERPAHPEAYNGLNLELNTHIAHAERNSPRVAVLTTRNEERFVAGLENLPH QGVQVYICNQSDRTLETACQYRGAGLIGESI PHNGWYQWQILRLRKBELFOSLEA DMLHIDAEIHLPSTYSLSVLSAADAALGNAIESSEFTFITREAPHDNDYQ HTLRTYPPRCPHFRAYKKQDGPETIAWSSGRVRFPEVPRFMRKGYILFL SAEHAARKYGGRRFSEEDYDGRGHWGRPLRPEDIQPDITQVTRTATDDDLDEANP WLAHWLRYAS"
complement(2648..3787)
/gene="ml17089"
complement(2648..3787)
/gene="ml17089"
/note="similar to mandelate racemase"
codon_start=1
/transl_table=11
/protein_id="BAB53263.1"
/db_xref="GI:14026667"
/translation="MSLSHFRITRFQASDRVIGDSQVADDDVNAALVELVSGEVG LGFTQLFPLPDQCQIESVFHEHVPKSLGNRAIALVHRVNRPGGQRAYSILPHE AVQVALMDLAKEAGLPLHLVLLGSRNRNRVAVASGLDFHLLDDDAFVSLFSHAASIGIS

AFKIKVGHDRDFDRDLRLRLKTCVPAGSKVMIDPNEAWTSKEALTKLVAIREAGHDL
LWVEDPILRHDDGLRTLTHAVTQINSGEYLDLQGRKLLLEAAADILNVHGOVTD
VWIKLAELGIPISIGNTELEAGVMAVALPEVEWLEYSEFONPHLVEQPIETRDG
YAYADPRPGHGLVSEKARGESRRRLARSLSGAAPENRPLPAK"
3870..4556
/gene="mlr7090"
3870..4556
/gene="mlr7090"
/codon_start=1
/transl_table=11
/product="transcriptional regulator"
/protein_id="BAB53264.1"
/db_xref="GI:14026668"
/translation="MAGKAKTDSPLVAARMSDIAYERILEGLFDRRVAGAFISQN
ELSAIVDPVAPLRDALRLEAEGLITIHPRSGIQVRPCELETATYQFRSIVERAA
VRVFAEQDEAVNMALEIRHSRLRRVERDQKCEQLEMDALEGLHGAIVAAALRNP
LIDSTVRMENVYLRLLLRDKITTPIMVRTLKEHLIDLEACGTRNADAAEAKLQAHFQ
AAMQRSGLGV"
4684..5922
/gene="mlr7091"
4684..5922
/gene="mlr7091"
/codon_start=1
/transl_table=11
/product="sugar ABC transporter, periplasmic binding protein"
/protein_id="BAB53265.1"
/db_xref="GI:14026669"
/translation="MQLNKRQFLMSAALAMAGHIRGARAASAISYWHHTTSOSEVAS
LLKTIIGLFOANADVVOTENINSEYMAKVSVAVLAGRPDPTWAIARFADLTAME
GLDLDLRVNGWKGKAFENRWITGLSSNGAIVAFAPADVMWYRADYFEBAGLAG
PPNPEFFISACRLTDPSKGRVAFQWRAGAGGFKYIVDMEAFGSPKIVKDGQPAIDR
AAAEIAITFYSDLEFLKEKVPVPSAPNDSYQINMEGRTQTMVWHHTGSLNLSAAM
KQKQFSTAMPAGPAKHAHVAVAYAGNIGMKEDNIEAARWVSFWGQTDAALLLSGT
GYFFSAALADMDPRITGNPIYHAAVETLEFGRLPNNFVGAAGWSENVPAPFAVLIG
QSTPEQAADRMIQGLEALR"
5979..6932
/gene="mlr7093"
5979..6932
/gene="mlr7093"
/codon_start=1
/transl_table=11
/product="ABC transporter, permealase protein"
/protein_id="BAB53266.1"
/db_xref="GI:14026670"
/translation="MLSHRGSMTSPSSGYRPLBALGELSETRYMVLILLPSLVMILL
VIAYPTLYGMAISFRENRLTRPGLGWVGFKHFAAMWSDPVFWIALKNTVWVTAAYV
LEALGLAATAALNRDLPGTKIFAVLILLPYFLPNVAVGHWALLDDPRLGVINDLLV
RAGVLOTYKAWFADPNTAATVAVAMHGFPPFALLLALGLKSI PEDLYKAAAVDGA
GVGAQLRLITPLMKLVIAAAVILRVIGLVNSPDLMLITGGGGGDTQVLSLYAFQT
AYKEFNFGYASALSVMVFMIFATAYIRLSRITKDGDD"
6925..7770
/gene="mlr7094"
6925..7770
/gene="mlr7094"
/codon_start=1
/transl_table=11
/product="ABC transporter, permealase protein"
/protein_id="BAB53267.1"
/db_xref="GI:14026671"
/translation="MIESKRRRVIANIATFYVLATLSMFCIGPMIMELTSLKYEAD
IVTQMTQYIPRRVTFVAVITQSGFPRLIANSVLVTFITVICATGILTAFAFR
PAFRGRTQLMGLYIVRMFVAVMMIIPLYVMRGVGLVDSRFGALAYTSLPLPLFVW
MLKGFDSAPKLESAARIDGSTRLGAMVRIVMPLARNGLVASSVFIAIAWNEFLFA
LMLTTQGSRTWPVGLQLMVGEFQLPWGLAAGSILSILPVILFAIVQRTMVGQLTA
GAIKG"
7781..8860
/gene="mlr7095"
7781..8860
/gene="mlr7095"
/codon_start=1
/transl_table=11
/product="sugar ABC transporter, ATP-binding protein"


```

QY 1435 GAAGCTTCGGCTGTGGAACAAAGTCTCTCTAATGTCGGTCTCTGCGCTTCGCGCTCG 1494
Db 338602 GACCGGTTGCGCGCTGGAAGTACGTCGTCTGATCGCTCTCTATATCTCGCGCTCGCG 338661
QY 1495 TTCTGTTCTTCATCGTGTATCTTTGTTCTATACGCTGTGCTCTGCGCGCGCGAGACCGCT 1554
Db 338662 GTGGCGATCTTCTCTACGGCGGTCTCGAAGCTTCAGAGAGCGCATCG---CCGCT 338718
QY 1555 CCGAACCGTGGCGGGAATTCGCGGATACGCTGTGATGAGCGTGCATCAGCGCTCGC 1614
Db 338719 GCCAACCATGGCGAGGCGGCACACACGCTGAAATGGCAGTGCCTTCGCGCGCGCG 338778
QY 1615 GCCACACGTTTCGAAACGCTGCCCAAGCGTCTGCACTGGGACAAAGCATCC 1664
Db 338779 TTCCACCAAGTGGAGCAGTTGCCGAAGATCAAGTAAGCGGCTCCCGCCC 338828

RESULT 13
LOCUS NWCOXABC 4380 bp DNA linear BCT 27-JUL-2001
DEFINITION N.winogradskyi DNA for coxA, coxB and coxC genes.
ACCESSION X89566
VERSION X89566.1 GI:1177572
KEYWORDS coxA gene; coxB gene; coxC gene; cytochrome c oxidase subunit I;
cytochrome c oxidase subunit II; Heme O synthase.
SOURCE Nitro bacter winogradskyi
ORGANISM Nitro bacter winogradskyi
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiaceae; Nitro bacter.
REFERENCE 1
AUTHORS Berben, G.
TITLE Nitro bacter winogradskyi cytochrome c oxidase genes are organized
in a repeated gene cluster
JOURNAL Antonie Van Leeuwenhoek 69 (4), 305-315 (1996)
MEDLINE 96433441
PUBMED 8836428
REFERENCE 2 (bases 1 to 4380)
AUTHORS Berben, G.P.R.
TITLE Direct Submission
JOURNAL Submitted (12-JUL-1995) G.P.R. Berben, Centre de Recherches
Agronomiques de l'Etat, Station de Chimie/Physique agric, Chaussee
de Wavre, 115, B-5030 Gembloux, BELGIUM
FEATURES
source
1. .4380
/organism="Nitro bacter winogradskyi"
/mol_type="genomic DNA"
/strain="ATCC 14123"
/db_xref="taxon:913"
/clone="plasmid Nw24.92"
/clone_lib="library of G.Berben"
727..1584
/gene="coxA"
727..1584
/gene="coxA"
/EC_number="1.9.3.1"
/codon_start=1
/transl_table=1
/product="cytochrome-c oxidase subunit II"
/protein_id="CAA61743.1"
/db_xref="GI:1177573"
/db_xref="GOA:Q59630"
/db_xref="TrEMBL:Q59630"
/translators="MMSMQYRRRFRPLGAFGGAFAAGAAFAADEIGQPHAW
GFDLQCVTFWMSIARMNGLLVITLITLVALLVAVYFRNARNPVSKTHN
TMIEVAWITLIPVILVGIAPVRLFEOLDLPKADLTATKATKQWNTYTPDNKFL
EPDLSLLAQDKKPLAVDNELVVPVNVKVRQVIGADVIHSFVPSFGMKVDIPEGL
NETWFKATKQGIYFGCGSELCLGRDHAFMPIAVRVSEQFAAWVETAKKKFASNPANA
VASAAGVQW"
727..828
/gene="coxA"
1682..3301
/gene="coxB"
1682..3301

```

```

/gene="coxB"
/EC_number="1.9.3.1"
/codon_start=1
/transl_table=1
/product="cytochrome-c oxidase subunit I"
/protein_id="CAA61744.1"
/db_xref="GI:1177574"
/db_xref="GOA:Q59631"
/db_xref="TrEMBL:Q59631"
/translators="MATAHADHDHGHHAHFTGWRKFAYSTNHKDIGMTYLVFA
IVAGITCGLSMGRIRAEMLYPGIOPDPAHTYVNFATNHLIMIIFMVMMPMIGGFGN
WFVPLMIGAPDMAPFRMNNISWILLPAAAFALLISFMBEGEPGACGATGWTLYAPLS
TSGHPGPAVDPSILSHLAGASSILGAINFITIENRAPGMTKMKPLFVMSILVTV
FLLALPVLGAGAITWLDNRNFTTFRADGGDPVLFQHLFWPFGPEVYVILLPG
FGMISHVSTFSKPIFGHLMAYANVAIGGIGFVVAHMYTGVKWSAQAYFVAAT
MWIATPGVKVFSIATMMGSLFEPMLFAIFLFTVGVTGCVLVANAGVORVM
QETTYVIAHPHYMGIAAFAIFSGYIYFWFPGYSGYMGNETTIGKLFHWMFPGANILF
PQHFGLQOMRMRYDYPDAFAGWNEISSYGAFTGFGVILIFLYGLVDFAFKQQA
DNWEGGATLTWLTSSPPPHQFSTLPKVQ"
3314..4303
/gene="coxC"
3314..4303
/gene="coxC"
/codon_start=1
/transl_table=1
/product="heme O synthase"
/protein_id="CAA61745.1"
/db_xref="GI:1177575"
/db_xref="GOA:Q51322"
/db_xref="TrEMBL:Q51322"
/translators="MCAHSCAHPKAGSIUSVVDKAVELPPRISEAGVADYFALL
KPVMSLVITLALVGIAPVHPVILGIALCIAVAGAGSALNMALEGDVIMS
RTANRPIPRITRGEAMGFLTSFVSMTGLVNYAGGLAFTIFFYVVIYTMG
LKRTAQNIYIGGAGALPVAWAAATGSLVEPLLLIIFFTPHFWALALFRS
DDYRAGVPMPLPVWAGPDATRLQILLYITVLVAIAAAPPLGYPFDWYIGVTLLIAG
MLVLAIEVYHRHTGSQALRATRLRFAFSIFYLFAVLVLLDWAQAPLIW"

ORIGIN
Query Match 42.4%; Score 709.2; DB 1; Length 4380;
Best Local Similarity 70.5%; Pred. No. 4.8e-137;
Matches 977; Conservative 0; Mismatches 403; Indels 6; Gaps 2;
QY 259 AACGTATGTTACCTACCATGTTCTCTGATGATGTTCTTTGTGGGTATCCCGCATG 318
Db 1910 AACGTGTTTCGACCAATCACGCGCTGATCATGATCTTCTCATGTGATGCTCGATG 1969
QY 319 TTGGTGTGTTTGGTAACATCTGATGCGCTGCAAAATCGGCGCTCGGATATGGCTTC 378
Db 1970 ATCGCGCGCTTCGCGCAACTGTTCTGTCGCGCTGATGATCGGCGCGCGACATGGCGTTC 2029
QY 379 CCGCGTATGAACAACCTGTCGTTCTGCGTGTTCATTGCGCGTACCGCGATGGCGTGGCT 438
Db 2030 CCGCGATGAACAACATCTGTTCTGCGTGTTCGCGCGCGCTTTGCGCTGCTGCTGATC 2089
QY 439 TCGCTGTTCCACCGCGCGGTGACGCTGAGTGGGTTCGGCGGTGGTGGTGGTTCGTGAC 498
Db 2090 TCTCGTTCATGGAGGTGAGCGCGCGCGACCGAGCGCGCGACCGGCTGGACGCTGTAT 2149
QY 499 CCGCGGTGTCGACCGCGCG---AAGTGTGCTATTTCGATGAGCTCGCGATTTTCGCGT 555
Db 2150 GCSCCGGTGTCGACGAGCGGTGATCCGGTCCGCGGTGAGCTTCAGCATCTGTCGATC 2209
QY 556 CACTTGTGGGTGCTCTCTGATCATGGCGGATCAACATGATCAGCATCTTTTGAC 615
Db 2210 CACTTGTGGGTGCTCTCTGATCATGGCGGATCAACTTCATCACCAGATCTTCAAC 2269
QY 616 ATGCGCGCGCGCGCATGACGCTGCACAAAGTGCCTGTTCTCTGTTGTCGATCTTTATC 675
Db 2270 ATGCGCGCGCTGCGATGACCATGACCAAGATGCCCTGTTCTGTTGTCGATCTTGTC 2329
QY 676 ACGGCTTGGCTGATCTCTGTCGGCTGCGGCTTCTGGCTGGTGCATCAATGCTGCTG 735
Db 2330 ACGGCTTCTGCTGCTGCGCTGCGGCTTCTGCGCGCGCGCATCAATGCTGCTG 2389

```

```

QY 736 ACCGACCGTAACCTCGGCACGACCTTCTTCAATCTGCTGGCGGGTACCCGATCTG 795
Db 2390 ACCGACCGTAACCTCGGCACGACGCTTCTCGCGCGGATGCGCGCGGATCCGGTCTG 2449
QY 796 TACCAACATCCTGTGTTCTTTGGGACCGGAAAGTGTACATCATCTTGCCTCCGGC 855
Db 2450 TTCCAGCATCTGTTCTGTTCTTGGTCAATCCGAGGTGTACATCTGATCTTCCCGGGC 2509
QY 856 TTGSCATCATCAGCCATGCTGTCGACCTTCTCGAAAGCGGCTTCTGCTTACCTG 915
Db 2510 TTGSCATCATCAGCCATGCTGTCGAGCTTCTCGAAGAGCCGATCTTCCGATATCTG 2569
QY 916 CGATGCTTATGMAATGTTGGCAATCGTGTCTGCGCTTCTGCGCTTCTGCGCGCACAC 975
Db 2570 GGCATGCGCTACGCCATGTTGGCATCGCGGCGCATCGGCTTCTGCTGTTGGCGCACAC 2629
QY 976 ATGTACACCGTTGTTGATGCTGCTGACCCAGCAATCTTACTTCTGCTGGCCACCATGGT 1035
Db 2630 ATGTACACCGTTGTTGATGCTGCTGCGGAGCGCTTACTTCTGCGCGGACCATGGT 2689
QY 1036 ATCCGCTGCGCAACCGCATTAAGATCTTCTGTTGATCGCCAGATGTGGCGGCTCG 1095
Db 2690 ATCCGCTGCGCAACCGCGTGAAGTGTCTGTTGATCGCCAGATGTGGCGGCTCG 2749
QY 1096 GTTCAAGTTCAATCGCGATGCTGCGGCTTCTGCTTATGTTCTGTTCAACGTTGGT 1155
Db 2750 ATCAAGTTCAAGCGCGATGCTGTTGCGATCGGCTTCTGTTCAACGTTGGT 2809
QY 1156 GGTGTACCGGTATGCTGCTGGCCAGCGGCTGACCGCTGATATCACGACATAT 1215
Db 2810 GGTGTACCGGTATGCTGCTGGCCAGTGTGCGGCTTCTGATCGCTGATGAGGAGCTAC 2869
QY 1216 TACGTGTGCGGACCTTCCATTAATGATGTCGTGGGTGCGATCTTTCGATCTTCGCG 1275
Db 2870 TACGTGTGCGGACCTTCCATTAATGATGTCGTGGGTGCGATCTTTCGATCTTCGCG 2929
QY 1276 GGTATCTACTTTTACATGCGGAGTCTCTGCGCGGCTTTCGCGGATGGGTGCAAG 1335
Db 2930 GGTATCTACTTTTACATGCGGAGTCTCTGCGGAGTCTCTGCGGATGGGTGCAAG 2989
QY 1336 CTGCACTCTGCGACCTTCTTCACTGCGGCAACGTCACGTTCTTCCGAGCACTTCCTG 1395
Db 2990 CTGCACTCTGCGGATGTTTCACTGCGGCAACATCTGTTCTTCCGAGCACTTCCTC 3049
QY 1396 GGACGTGAGGTATGCGCGCGCTTACATGATGATGCGGAGCTTCCGAGCTTCCGCTGGAC 1455
Db 3050 GGCTGAGGCGATGATGCGGCGATGTTGAGATATCCGATGCTTCCGCGGCTGGAAC 3109
QY 1456 AAGTCTGCTATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1515
Db 3110 GAGATCTGCTTCTACGCGGCTTCTACGCGGCTTCCGCGTCTATCTTCTTCTACGCG 3169
QY 1516 TTGTCTATACGCTGCTGCTGCGCGCGGAGACCGCTTCCGAGCCGTTGGGCGAATC 1575
Db 3170 CTGCTGAGCGC---CTTCCGCGCAAGCAGCAGCGGCTGCAATCTCGGCGTGAAGC 3226
QY 1576 GCGGATACGCTGGAATGAGCGCTGCGCATCACCGCTTCCGCGCGCACACGTTTGAACGCTG 1635
Db 3227 GCGGAGCGTGAATGAGCGCTGCTTCTCACCGCTTCCGCTTCCACAGTTCTTCCACCTG 3286
QY 1636 CCCAAG 1641
Db 3287 CCGAAG 3292

```

RESULT 14

```

LOCUS      BJCOXAG
DEFINITION B.japonicum coxA gene for cytochrome aa(3) subunit I (EC 1.9.3.1).
ACCESSION X54318.1
VERSION    X54318.1 GI:39505
KEYWORDS   coxA gene; cytochrome aa(3) subunit I.

```

SOURCE

Bradyrhizobium japonicum
Bradyrhizobium japonicum
Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
Bradyrhizobiaceae; Bradyrhizobium.

REFERENCE

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

JOURNAL

91045095
2172930
2 (bases 1 to 1830)

PUBMED

Gabel, C.
Direct Submission
Submitted (07-AUG-1990) Gabel C., Johns Hopkins University, Dept.
of Biology, Charles and 34th Sts., Baltimore, MD 21218, USA

AUTHORS

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

JOURNAL

91045095
2172930
2 (bases 1 to 1830)

FEATURES

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

source

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

RBS

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

CDS

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

ORIGIN

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

Query Match

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

Best Local Similarity

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

Matches

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

Conservative

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

Mismatches

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

Indels

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

Gaps

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

Score

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

DB 1

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

Length

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

DB 1

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

DB 1

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

DB 1

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

DB 1

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

DB 1

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

DB 1

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

DB 1

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

DB 1

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

DB 1

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

DB 1

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

DB 1

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

DB 1

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

DB 1

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

DB 1

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

DB 1

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

DB 1

1
Gabel, C. and Maier, R. J.
Nucleotide sequence of the coxA gene encoding subunit I of
cytochrome aa3 of Bradyrhizobium japonicum
Nucleic Acids Res. 18 (20), 6143 (1990)

607	QY	TTCTTTGAACATGCGCGCCCGCGCATGACGCTGCACAAAGTGCCTGTGTTCTCTGTGGTCTG	666
763	Db	ATCTTTCAACATGCGCGCGCGCGGCATGACCCCTGSCAAAGATGCCTGTGTTCTGTGTGGTCTG	822
667	QY	ATCTTTATCACGGGTTTGGCTGTATCTCTGTGGCGCTCCGGTTCTGGCTGTGGTGAACATCAC	726
823	Db	ATCTGTGTGACCGTGTTCTCTGTCTGTGTGTGCTGCGCTGCTGCTCGCGGTGCGATCAC	882
727	QY	ATGCTGTGTGACCGACCGTACTTTCGCGACGACCTTCTCAATCTGCTGCGCGCGCTGAC	786
883	Db	ATGCTGTCTACCGACCGCAATTTTGGCAGCAGCTTCTTCGCGCCGACGCGCGCGCGGAC	942
787	QY	CCGATTCTGTACCAACACATCTCTGTGGTTCTTTTGGGCACCCGGAAGTGTATCATCATTT	846
943	Db	CCCGTGTGTCTCAGCACCTGTCTTGGTTCTTTCGGCCATCCGAGGTGTATCATCTGATC	1002
847	QY	CTGCGCGGCTTTGGCATCATCAGCCATCTGCTGCTGACCTTCTCGAAAGTGCCTGCTTC	906
1003	Db	CTGCGCGGCTTCGCGATGATCAGCCAGATCTGCTCCACCTTCTCGCGAAGCCCGTGTTC	1062
907	QY	GGTTACCTGCCATGTGTTCTATGCAATGTGGCAATCGGTGTTCTGGGCTTTGTCTGTG	966
1063	Db	GGCTATCTCGGCATGGCTACGCCATGTGTCGGCATCGCGCATCGGCTCGTGGTGTG	1122
967	QY	CGCGACACATGTACACCGTTGGTATGTCGCTGACCCAGCAATCCTACTTCATCTGCTGGCC	1026
1123	Db	CGGACACATGTACACGTCGCGATCTGCTCGGCGACGCGAGCTATTTCTGTCGCGCG	1182
1027	QY	ACCATGTGTATCGCGTGTCCGACCCGCCATTAAGAATCTTCTCTGTGATGCGCACGATGTGG	1086
1183	Db	ACGATGTGTATCGCGTTCGCGACCGCGCTGGAAGATCTTCTGTGATCGCGACGATGTGG	1242
1087	QY	GGCGGCTCGGTTGAGTTTCAAAATCGCGCATGTCTCGGGCTTTTGGCTTTATGTTCTCTGTTTC	1146
1243	Db	GGCGGCTCGATCGAATTTCCGCGCGCGCATGATCTCGGCGCGTGGGTCTATCTTCTGTTTC	1302
1147	QY	ACCGTGGGTGTGTGACCGGTATGTTGTGGCCCCAAGCGGGTGTGGACCGTGCAATACAC	1206
1303	Db	ACGGTCTGGTGGCTTACCGCGCTGTGCTGTGGCGAAACCGCGGCTCGACCGCTGCTCCAG	1362
1207	QY	GACACCTATTACGTTGGTGGCGCACTTCCATTATGTGATGTCTGCTGGGTGGATCTTTGG	1266
1363	Db	GAGACCTACTACGTCGTGCGCGCACTTCCACTACGTGCTGTGCTCTGCTGCGATGTTCTCG	1422
1267	QY	ATCTTCCCGGTATPCTACTTTTACATGCGGAAGTTCTCGGCGCGGCTTCCCGGAATGG	1326
1423	Db	ATCTTCCCGGCTGGTACTACTGGTTTCCGGAAGATGACGGGTCATATGTACACGAGACG	1482
1327	QY	GCTCCAAAGCTGCACCTCTGGACCTTCTCATCGGTGGGACGTCACGTTCTTCTCCCGCAG	1386
1483	Db	CTGGCGAAGCGGCATCTTCTGGGTCACTTCACTCGGGGTGAACCTGTGTCTTCTCCCGCAG	1542
1387	QY	CACCTCTCTGGAAGTCAAGGTATCCGCGCGCTTACATCGACTATCCGAGAGCTTTCGG	1446
1543	Db	CACCTCTCTGCGCTGTCCGGCATCGCGCGCGCTACGTCGACTATCCGACGCGTTCGCC	1602
1447	QY	CTGTGGAAACAAGTCTCGTCTATGTGTGCTTCTTGGCTTCGCTCGCTCTCTGTTCTTTC	1506
1603	Db	GGCTGGAAACCTGTGTCGTCTGCTGGCTCTTACATTTTCGGGTTCGGGCTCTGATCTTC	1562
1507	QY	ATCGTGTATCTTTGTCTATACGCTGTGTGTGCGCGCGAGACCGCTCGAACCCTGTGG	1566
1663	Db	CTCTATTTCGCTGTATCGATGC---CTTTGGGAAAAGAGTGCCTGGTGGCGCAATTCCTGTGG	1719
1567	QY	GGCGAATTCGCGGATACGCTGGAAATGGACGCTGGCAATCACCGCTCTCGGCGCCACCGTTC	1626
1720	Db	GGTCCGCGCGACACGCTGGAGTGTGAAGCTGCTCCCTCGCGCGCCCTTCCATCAGTTC	1779
1627	QY	GAACGCTTGCCC	1638
1780	Db	GAAGTGTGCCC	1791

```

RESULT 15
BUCOXA          2000 bp      DNA      linear      BCT 07-NOV-1991
B-japonicum coxa gene for subunit I of cytochrome c oxidase
(SC_number=1.9.3.1).
ACCESSION      X54800
VERSION        X54800.1 GI:39502
KEYWORDS       cytochrome c oxidase.
SOURCE         Bradyrhizobium japonicum
ORGANISM       Bradyrhizobium japonicum
REFERENCE      1 (bases 1 to 2000)
AUTHORS        Bott,M., Bolliger,M. and Hennecke,H.
TITLE          Genetic analysis of the cytochrome c-a3 branch of the
                Bradyrhizobium japonicum respiratory chain
JOURNAL        Mol. Microbiol. 4 (12), 2147-2157 (1990)
MEDLINE        92211625
PUBMED         1965217
REFERENCE      2 (bases 1 to 2000)
AUTHORS        Bott,M.
TITLE          Direct Submission
JOURNAL        Submitted (10-OCT-1990) Bott M., Eidgenossische Technische
                Hochschule, Mikrobiologisches Institut, Schmelzbergstrasse 7,
                CH-8092 Zurich, Switzerland
FEATURES       source
                1..2000
                /organism="Bradyrhizobium japonicum"
                /mol_type="genomic DNA"
                /strain="USDA311b110spc4"
                /db_xref="taxon:375"
                /clone="PRJ3404"
                75..1700
                /gene="coxA"
                75..1700
                /gene="coxA"
                /codon_start=1
                /transl_table=11
                /product="cytochrome-c oxidase"
                /protein_id="CAA38570.1"
                /db_xref="GI:39503"
                /db_xref="GOA:P31833"
                /db_xref="Swiss-Prot:P31833"
                /translation="WATSAARHGDAQHQHDEHAHGTGRRVYVSTNKHDKICTMYLI
                GNVIAGVIAAMSIARAEIMYPGQIIFETHTYNVFTVSHGLIMIFFWMPANIGGF
                FAWFVIMTGADPAPRMNNISFWLJLPASFGLLLMSTFVEGEFGANGAGWMTVP
                LSSSGHPAVDPALISLHAGSSILGANFITTENRAPQMTLHKXELFVMSILV
                TVFLLLSLPVLAGAITMLLTDNRNFTTFAPDGGSDVLFQHLFWFPGHPEVYIIL
                PGMSTQSIVTFSKPPVGLGWAYMAVIGGIFVVAWHMYTVGMSATQAVFA
                ATNVIAPGVKIFSWIATMGGSTEFAPMIMAVGFILFTYGGVTGVVLNAGVDR
                VIGETYYVAHHYVLSLGAFAIPAGWYWFPMXTGYMNEHLAKAHFWFTFIGNL
                VFPPFGLSGMRRYVDYDFAPAGNVLSSVSGVYISGFGVLIIFYCVIDAPAKKVP
                AGNPNMGAGATLEWTLPSPPPPHQFVILPQV"
                75..1697
                /gene="coxA"
                /product="cytochrome-c oxidase"
                1854..>2000
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA38571.1"
                /db_xref="GI:39504"
                /db_xref="GOA:Q99253"
                /db_xref="TrEMBL:Q99253"
                /translation="NSLVIFTLVGMAMAPGHPPHVLAIITSLLCIAVGAGASCALNNA
                LEGDI"
                1854..>2000
                /product="unnamed"
mat_peptide
                75..1697
                /product="cytochrome-c oxidase"
                1854..>2000
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA38571.1"
                /db_xref="GI:39504"
                /db_xref="GOA:Q99253"
                /db_xref="TrEMBL:Q99253"
                /translation="NSLVIFTLVGMAMAPGHPPHVLAIITSLLCIAVGAGASCALNNA
                LEGDI"
                1854..>2000
                /product="unnamed"
mat_peptide
                75..1697
                /product="cytochrome-c oxidase"
                1854..>2000
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA38571.1"
                /db_xref="GI:39504"
                /db_xref="GOA:Q99253"
                /db_xref="TrEMBL:Q99253"
                /translation="NSLVIFTLVGMAMAPGHPPHVLAIITSLLCIAVGAGASCALNNA
                LEGDI"
                1854..>2000
                /product="unnamed"
mat_peptide
                75..1697
                /product="cytochrome-c oxidase"
                1854..>2000
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA38571.1"
                /db_xref="GI:39504"
                /db_xref="GOA:Q99253"
                /db_xref="TrEMBL:Q99253"
                /translation="NSLVIFTLVGMAMAPGHPPHVLAIITSLLCIAVGAGASCALNNA
                LEGDI"
                1854..>2000
                /product="unnamed"
mat_peptide
                75..1697
                /product="cytochrome-c oxidase"
                1854..>2000
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA38571.1"
                /db_xref="GI:39504"
                /db_xref="GOA:Q99253"
                /db_xref="TrEMBL:Q99253"
                /translation="NSLVIFTLVGMAMAPGHPPHVLAIITSLLCIAVGAGASCALNNA
                LEGDI"
                1854..>2000
                /product="unnamed"
mat_peptide
                75..1697
                /product="cytochrome-c oxidase"
                1854..>2000
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA38571.1"
                /db_xref="GI:39504"
                /db_xref="GOA:Q99253"
                /db_xref="TrEMBL:Q99253"
                /translation="NSLVIFTLVGMAMAPGHPPHVLAIITSLLCIAVGAGASCALNNA
                LEGDI"
                1854..>2000
                /product="unnamed"
mat_peptide
                75..1697
                /product="cytochrome-c oxidase"
                1854..>2000
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA38571.1"
                /db_xref="GI:39504"
                /db_xref="GOA:Q99253"
                /db_xref="TrEMBL:Q99253"
                /translation="NSLVIFTLVGMAMAPGHPPHVLAIITSLLCIAVGAGASCALNNA
                LEGDI"
                1854..>2000
                /product="unnamed"
mat_peptide
                75..1697
                /product="cytochrome-c oxidase"
                1854..>2000
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA38571.1"
                /db_xref="GI:39504"
                /db_xref="GOA:Q99253"
                /db_xref="TrEMBL:Q99253"
                /translation="NSLVIFTLVGMAMAPGHPPHVLAIITSLLCIAVGAGASCALNNA
                LEGDI"
                1854..>2000
                /product="unnamed"
mat_peptide
                75..1697
                /product="cytochrome-c oxidase"
                1854..>2000
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA38571.1"
                /db_xref="GI:39504"
                /db_xref="GOA:Q99253"
                /db_xref="TrEMBL:Q99253"
                /translation="NSLVIFTLVGMAMAPGHPPHVLAIITSLLCIAVGAGASCALNNA
                LEGDI"
                1854..>2000
                /product="unnamed"
mat_peptide
                75..1697
                /product="cytochrome-c oxidase"
                1854..>2000
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA38571.1"
                /db_xref="GI:39504"
                /db_xref="GOA:Q99253"
                /db_xref="TrEMBL:Q99253"
                /translation="NSLVIFTLVGMAMAPGHPPHVLAIITSLLCIAVGAGASCALNNA
                LEGDI"
                1854..>2000
                /product="unnamed"
mat_peptide
                75..1697
                /product="cytochrome-c oxidase"
                1854..>2000
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA38571.1"
                /db_xref="GI:39504"
                /db_xref="GOA:Q99253"
                /db_xref="TrEMBL:Q99253"
                /translation="NSLVIFTLVGMAMAPGHPPHVLAIITSLLCIAVGAGASCALNNA
                LEGDI"
                1854..>2000
                /product="unnamed"
mat_peptide
                75..1697
                /product="cytochrome-c oxidase"
                1854..>2000
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA38571.1"
                /db_xref="GI:39504"
                /db_xref="GOA:Q99253"
                /db_xref="TrEMBL:Q99253"
                /translation="NSLVIFTLVGMAMAPGHPPHVLAIITSLLCIAVGAGASCALNNA
                LEGDI"
                1854..>2000
                /product="unnamed"
mat_peptide
                75..1697
                /product="cytochrome-c oxidase"
                1854..>2000
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA38571.1"
                /db_xref="GI:39504"
                /db_xref="GOA:Q99253"
                /db_xref="TrEMBL:Q99253"
                /translation="NSLVIFTLVGMAMAPGHPPHVLAIITSLLCIAVGAGASCALNNA
                LEGDI"
                1854..>2000
                /product="unnamed"
mat_peptide
                75..1697
                /product="cytochrome-c oxidase"
                1854..>2000
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA38571.1"
                /db_xref="GI:39504"
                /db_xref="GOA:Q99253"
                /db_xref="TrEMBL:Q99253"
                /translation="NSLVIFTLVGMAMAPGHPPHVLAIITSLLCIAVGAGASCALNNA
                LEGDI"
                1854..>2000
                /product="unnamed"
mat_peptide
                75..1697
                /product="cytochrome-c oxidase"
                1854..>2000
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA38571.1"
                /db_xref="GI:39504"
                /db_xref="GOA:Q99253"
                /db_xref="TrEMBL:Q99253"
                /translation="NSLVIFTLVGMAMAPGHPPHVLAIITSLLCIAVGAGASCALNNA
                LEGDI"
                1854..>2000
                /product="unnamed"
mat_peptide
                75..1697
                /product="cytochrome-c oxidase"
                1854..>2000
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA38571.1"
                /db_xref="GI:39504"
                /db_xref="GOA:Q99253"
                /db_xref="TrEMBL:Q99253"
                /translation="NSLVIFTLVGMAMAPGHPPHVLAIITSLLCIAVGAGASCALNNA
                LEGDI"
                1854..>2000
                /product="unnamed"
mat_peptide
                75..1697
                /product="cytochrome-c oxidase"
                1854..>2000
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAA38571.1"
                /db_xref="GI:39504"
                /db_xref="GOA:Q99253"
                /db_xref="TrEMBL:Q99253"
                /translation="NSLVIFTLVGMAMAPGHPPHVLAIITSLLCIAVGAGASCALNNA
                LEGDI"
                1854..>2000
                /product="unnamed"
mat_peptide
                75..1697
                /product="cytochrome-c oxidase"
                1
```

ORIGIN

Query Match	42.3%;	Score 708.8;	DB 1;	Length 2000;
Best Local Similarity	70.3%;	Pred. No. 6.1e-137;		
Matches 979: Conservative	0;	Mismatches 407;	Indels 6;	Gaps 2;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 22, 2004, 12:20:45 ; Search time 996 Seconds
(without alignments)

8822.822 Million cell updates/sec

Title: US-09-712-768A-1

Perfect score: 1674

Sequence: 1 atggcagagccgccattca.....acaagcatccctcgcaactaa 1674

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23Sep04.*

1: Geneseqn1980s.*
2: Geneseqn1990s.*
3: Geneseqn2000s.*
4: Geneseqn2001as.*
5: Geneseqn2001bs.*
6: Geneseqn2002as.*
7: Geneseqn2002bs.*
8: Geneseqn2003as.*
9: Geneseqn2003bs.*
10: Geneseqn2003cs.*
11: Geneseqn2003ds.*
12: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1674	100.0	1674	4	AAA91489 G. oxydan
2	515.6	30.8	1593	8	ABT14590 Pseudomon
3	503.6	30.1	1359	11	ABD02101 Pseudomon
4	503.6	30.1	1587	11	ABD02072 Pseudomon
5	489	29.2	1278	11	ABD02134 Pseudomon
6	408.4	24.4	16568	8	ACC45136 Human mit
7	406.2	24.3	1629	8	ACA37676 Prokaryot
8	405	24.2	1555	5	AAS75523 DNA encod
9	398.8	23.8	16559	10	ADD42976 Human mit
10	398.8	23.8	16561	10	ADD42971 Human mit
11	398.8	23.8	16568	10	ADD43240 Human mit
12	398.8	23.8	16569	10	ADD43266 Human mit
13	398.8	23.8	16569	10	ADD43327 Human mit
14	398.8	23.8	16569	10	ADD43295 Human mit
15	398.8	23.8	16569	10	ADD43302 Human mit
16	398.8	23.8	16569	10	ADD43334 Human mit
17	398.8	23.8	16569	10	ADD43355 Human mit
18	398.8	23.8	16569	10	ADD43350 Human mit
19	398.8	23.8	16569	10	ADD43007 Human mit
20	398.8	23.8	16569	10	ADD43300 Human mit
21	398.8	23.8	16570	10	ADD42959 Human mit

ALIGNMENTS

RESULT 1

AAA91489

ID AAA91489 standard; DNA; 1674 BP.

XX AC AAA91489;

XX AC AAA91489;

XX DT 06-AUG-2001 (first entry)

XX DE G. oxydans cytochrome C oxidase (COI) coding sequence.

XX KW Cytochrome C oxidase; COI; COII; COIII; COI complex; D-sorbitol;

XX KW oxidative fermentation; electron transfer; respiratory chain; L-sorbose;

XX KW 2-Keto-L-gluconic acid; 2KGA production; aldehyde production;

XX KW carboxylic acid production; ketone production; ds.

XX OS Gluconobacter oxydans.

XX FH Key Location/Qualifiers

XX FT CDS 1..1674

XX FT /*tag= a

XX FT /product= "COI"

XX PN EPI103603-A2.

XX PD 30-MAY-2001.

XX PF 14-NOV-2000; 2000EP-00124785.

XX PR 17-NOV-1999; 99EP-00122842.

XX PA (HOFF) HOFFMANN LA ROCHE & CO AG F.

XX PI Asakura A, Hoshino T, Shinjoh M;

XX DR WPI; 2001-357953/38.

XX DR P-ESDB; AAY97750.

XX PT New cytochrome c oxidase complex having cytochrome c oxidase activity from Gluconobacter oxydans DSM 4025, useful in mediating electron transfer in respiratory chain or producing 2-keto-L-gulonic acid from L-sorbose or D-sorbitol.

XX PS Claim 13; Page 16-20; 42pp; English.

XX XX

CC This sequence encodes the Gluconobacter oxydans cytochrome C oxidase
 CC (COI) of the invention. The COI complex is useful in improving oxidative
 CC fermentation and is an essential component mediating electron transfer in
 CC the respiratory chain. The recombinant microorganism and the cytochrome c
 CC oxidase may be used in the genetic preparation of a recombinant COI
 CC complex and in the production of 2-Keto-L-gluconic acid (2KGA) from L-
 CC sorbose or D-sorbitol in a culture medium. The COI is also useful as a
 CC terminal oxidase, oxidising cytochrome c, an electron acceptor from an
 CC enzyme belonging to dehydrogenase for the production of aldehydes,
 CC carboxylic acids and ketones from alcohols and aldehydes, especially the
 CC production of 2KGA from L-sorbose or D-sorbitol
 XX
 SQ Sequence 1674 BP; 270 A; 493 C; 468 G; 443 T; 0 U; 0 Other;

Query Match 100.0%; Score 1674; DB 4; Length 1674;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1674; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCAGACCGCGCCATTACAGCCCATGACACCATGAGAGCAAGGCTTCTTCAACGCGC 60
 DB 1 ATGGCAGACCGCGCCATTACAGCCCATGACACCATGAGAGCAAGGCTTCTTCAACGCGC 60

QY 61 TGGTTTCATGTCGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 120
 DB 61 TGGTTTCATGTCGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 120

QY 121 GTTGGTTTCATGTCGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 180
 DB 121 GTTGGTTTCATGTCGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 180

QY 181 GTTCAGTACATGTCGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 240
 DB 181 GTTCAGTACATGTCGACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 240

QY 241 GCGAACGACACCTGTGAAAGCTCATGTTTACCTACCATGATGTTTCTGATGATGTTTCTTT 300
 DB 241 GCGAACGACACCTGTGAAAGCTCATGTTTACCTACCATGATGTTTCTGATGATGTTTCTTT 300

QY 301 GTGGTATCCCGCATTTGTCGGTGGTTTGGTAACTATCTGATGCGCGTGCAGAAATCGGC 360
 DB 301 GTGGTATCCCGCATTTGTCGGTGGTTTGGTAACTATCTGATGCGCGTGCAGAAATCGGC 360

QY 361 GCTCCGGATATGGCTTCCCGCATGAAACCACTGCTGTTCTGGCTGTTCAITGCGCGGT 420
 DB 361 GCTCCGGATATGGCTTCCCGCATGAAACCACTGCTGTTCTGGCTGTTCAITGCGCGGT 420

QY 421 ACCGGATGGCGGTGGCTTCCCGCATGAAACCACTGCTGTTCTGGCTGTTCAITGCGCGGT 480
 DB 421 ACCGGATGGCGGTGGCTTCCCGCATGAAACCACTGCTGTTCTGGCTGTTCAITGCGCGGT 480

QY 481 GTTGGTTGGTCTGTACCCCGCGCTGTGACCCCGCATGACCTGATTCGATGGACCTTC 540
 DB 481 GTTGGTTGGTCTGTACCCCGCGCTGTGACCCCGCATGACCTGATTCGATGGACCTTC 540

QY 541 GCGATTTTCGGGTTTCACTTGTCCGGTCCCTGCTGATGCGCGCATCAACATGATC 600
 DB 541 GCGATTTTCGGGTTTCACTTGTCCGGTCCCTGCTGATGCGCGCATCAACATGATC 600

QY 601 ACGACCTTCTTGAACATCGCGCGCCCGCGCATGACCTGTCACAAAGTGCCTGTTGTTCTCG 660
 DB 601 ACGACCTTCTTGAACATCGCGCGCCCGCGCATGACCTGTCACAAAGTGCCTGTTGTTCTCG 660

QY 661 TGGTGCATCTTATCAGCGCTTGGGTATCTGCTGGCGGTGCGGTGCTGGCTGGTGA 720
 DB 661 TGGTGCATCTTATCAGCGCTTGGGTATCTGCTGGCGGTGCGGTGCTGGCTGGTGA 720

QY 721 ATCACCATGCTGCTGACCGACCGTAACCTTCGGACGACCTTCTTCAATCTGCTGGCGGC 780
 DB 721 ATCACCATGCTGCTGACCGACCGTAACCTTCGGACGACCTTCTTCAATCTGCTGGCGGC 780

QY 781 GGTGACCGGATCTGTACCAACACATCTCTGTTCTTGGGACCCCGGAAGTGTACATC 840
 DB 781 GGTGACCGGATCTGTACCAACACATCTCTGTTCTTGGGACCCCGGAAGTGTACATC 840

QY 841 ATCATCTGCGCGCTTTGGCATCATCAGCATATGCTGTCGACCTTCTCGAAAAAGCGG 900
 DB 841 ATCATCTGCGCGCTTTGGCATCATCAGCATATGCTGTCGACCTTCTCGAAAAAGCGG 900

QY 901 GTCTTCGTTACCTGCGCATGCTTATGCAATGCTGCAATCGTGTCTTCTGGGCTTTGTC 960
 DB 901 GTCTTCGTTACCTGCGCATGCTTATGCAATGCTGCAATCGTGTCTTCTGGGCTTTGTC 960

QY 961 GTCTGGGCGCACCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020
 DB 961 GTCTGGGCGCACCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020

QY 1021 CTGGCCACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080
 DB 1021 CTGGCCACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080

QY 1081 ATGTCGGGCGCTCGGTTGAGTTCAAAATCGCCGATGCTCTGGGCCCTTGGCTTTATGTTTC 1140
 DB 1081 ATGTCGGGCGCTCGGTTGAGTTCAAAATCGCCGATGCTCTGGGCCCTTGGCTTTATGTTTC 1140

QY 1141 CTGTTTCCCGTGGTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200
 DB 1141 CTGTTTCCCGTGGTGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200

QY 1201 TATCAGACACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260
 DB 1201 TATCAGACACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1260

QY 1261 TTTGGGATCTTCGCGCGTACTACTTTTACATCCGAGTTCGCGGCGCGCTTCCCG 1320
 DB 1261 TTTGGGATCTTCGCGCGTACTACTTTTACATCCGAGTTCGCGGCGCGCTTCCCG 1320

QY 1321 GAATGGGCTGCAAAAGCTGCACTTCTGGACCTTCTTATCGGTGCGAACCTGCTGCTTCTTC 1380
 DB 1321 GAATGGGCTGCAAAAGCTGCACTTCTGGACCTTCTTATCGGTGCGAACCTGCTGCTTCTTC 1380

QY 1381 CCGCAGACACTTCTCGGACCTGAGGTATGCGCGCGCTTACATGACATATCCCGAAGCC 1440
 DB 1381 CCGCAGACACTTCTCGGACCTGAGGTATGCGCGCGCTTACATGACATATCCCGAAGCC 1440

QY 1441 TCGCGCTGTGGACAAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
 DB 1441 TCGCGCTGTGGACAAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500

QY 1501 TTTCTCATCGTGTCTTGTCTATACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560
 DB 1501 TTTCTCATCGTGTCTTGTCTATACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1560

QY 1561 CCGTGGGCGCAATTCGCGCATGCGTGGAAATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
 DB 1561 CCGTGGGCGCAATTCGCGCATGCGTGGAAATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620

QY 1621 ACGTTCGAACGCTGCGCAAGCGCTGCGACTGCGACAAAGCATCCCTCGCACTAA 1674
 DB 1621 ACGTTCGAACGCTGCGCAAGCGCTGCGACTGCGACAAAGCATCCCTCGCACTAA 1674

RESULT 2

ABT14590

ID ABT14590 standard; DNA; 1593 BP.

XX ABT14590;

XX 27-FEB-2003 (first entry)

XX Pseudomonas aeruginosa biofilm formation-related gene #32.

XX Gene; ds; biofilm formation modulation; biofilm-associated disease;

XX cystic fibrosis; AIDS; middle ear infection; acne; periodontal disease;

XX catheter-associated infection; medical device-associated infection.

XX Pseudomonas aeruginosa.

XX OS Pseudomonas aeruginosa.
XX FN US6551795-B1.
XX FD 22-APR-2003.
XX PF 18-FEB-1999; 99US-00252991.
XX PR 18-FEB-1998; 98US-0074788P.
XX PR 27-JUL-1998; 98US-0094190P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX DR WPI; 2003-615309/58.
XX DR P-PSDB; ABO68530.
XX PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
XX PT useful as molecular targets for diagnostics, prophylaxis and treatment of
XX PT pathological conditions resulting from bacterial infection.
XX PS Disclosure; SEQ ID NO 705; 455pp; English.
XX CC The invention relates to Pseudomonas aeruginosa polypeptides and the
XX CC polynucleotides encoding them. The sequences are useful in diagnosis and
XX CC therapy of pathological conditions, as molecular targets for diagnostics,
XX CC prophylaxis and treatment of pathological conditions resulting from a
XX CC bacterial infection, for evaluating a compound, such as a polypeptide,
XX CC for the ability to bind a P. aeruginosa nucleic acid, as components of
XX CC effective antibacterial targets, as targets for antibacterial drugs,
XX CC including anti-P. aeruginosa drugs, as templates for recombinant
XX CC production of P. aeruginosa-derived peptides or polypeptides, as target
XX CC components for diagnosis and/or treatment of P. aeruginosa-caused
XX CC infection, and in detection of P. aeruginosa sequences or other sequences
XX CC of Pseudomonas species using biochip technology. Sequences ABD01397-
XX CC ABD1967 represent P. aeruginosa polynucleotides of the invention. Note:
XX CC The sequence data for this patent did not form part of the printed
XX CC specification but was obtained in electronic format from USPTO at
XX CC seqdata.uspto.gov/sequence.html
XX SQ Sequence 1359 BP; 193 A; 487 G; 389 G; 290 T; 0 U; 0 Other;

Query Match 30.1%; Score 503.6; DB 11; Length 1359;
Best Local Similarity 63.5%; Pred. No. 1.3e-124;
Matches 841; Conservative 0; Mismatches 464; Indels 19; Gaps 4;

QY 319 TTCGGTGGTTTGGTAATCTGATGCGCGCTGCAATCGCGCTCCGGATATGG-CCTT 377
DB 31 TTGCTCGGCTGCCAACTGGATGATCCCGCTGATGATCGGCGCGCGACATGCCCGCT 90

QY 378 CCGCGTATGAACAACCTGTGCTTGGCTGTTCATGCGGTACCGGATGGGGGTGGC 437
DB 91 GCGCGGATGAACAATTCAGCTTCTGGCTGTGCTGCGGGGCTTCGGCTCTGCTCAG 150

QY 438 TTGCGTGTTCGACACGGCGGTGACGGTACGTGGTTCGGCGGTGTGGTTCGTGTA 497
DB 151 CACCTGTGTTCATCGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 198

QY 498 CCGCGCGCTGTGACACCGCGGAACTGGCTATTTCATGAGACCTCGCGATTTTCGGGTTCA 557
DB 199 TCGCGCGCTGTGACACACCTTCGCGCGGACAGCTGACCTTCATCTTCGCGATCCA 258

QY 558 CTTGTGGGTGCTCTCGATCATGGCGCGATCAACATGATCAGACCTCTTGAACAT 617
DB 259 CTTGGCGGGATCAGCTCGATCATGGCGCGATCAACGTGATCGCCACCATCTCAACT 318

QY 618 GCGCGCGCGCGCGATGACCTGCACAAGTGGCTGTGTTCTCGTGGTGCATCTTATCAC 677
DB 319 GCGCGCGCGCGCGATGACCTTCATGAGATGCGGTGTTCGTCTGGACCTGGCTGATCAC 378

QY 678 GGTGGTGGTGTGCTGCTGGCGCTCCCGGTTCCTGGCTGGTGAATCACCATGCTGCTGAC 737

RESULT 4
ABD02072/c
ID ABD02072 standard; DNA; 1587 BP.
XX
AC ABD02072;

DB 379 CCGCTTCCTGCTGATCGCGGTGATCGCGGTGCTGCGCGGTGCTGATGATGCTGAT 438
QY 738 CGACCGTAACCTCGGACGACCTTCTTCAATCCTGCTGGCGGCGGTGACCGATCTGTGA 797
DB 439 GGACATCCACTTCGGCACCACTTCTTTCAGCGCGCGCGCGCGCGCGCGCGGTGCT 498
QY 798 CCAACACATCCTGTGGTTCCTTGGGCAACCGGAAAGTGTACATCATCATTTGCGCGGCTT 857
DB 499 CCAGCACGTGTTCTGGTTCCTTTCGGCCACCCCGAGGTGTACATCATCATCTCTGCGCGCTT 558
QY 858 TGGCATCATCAGCCATGTCGTGCGACCTTCTGAAAGCGCGGTCTTCGCTTACCTGCC 917
DB 559 CGGTGGGTGATGCGGATCATCCGACCTTCGCGCGCAAGCGCTGTTCGCGTACACCTC 618
QY 918 GATGCTCTATGCAATGTGGCAATCGGTGTTCTGGGCTTTGTCGTCTGGGCGCACCAT 977
DB 619 GATGCTCTAGCCACCGCCAGCATCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 678
QY 978 GTACACGTTGGTATGCTGACCGACCAATCCTACTTATGCTGCGCACCATGCTGAT 1037
DB 679 GTTCGTGTTGGCATCCCGGTACCGCGGAGCTTCTTCTCATGTACGCCACCATGCTGAT 738
QY 1038 CCGGTGCGGACCGGCATTAAGATCTTCTGCTGATCGCCACGATGTGGGCGGCTCGGT 1097
DB 739 CCGGTGCGGACCGGCATTAAGATCTTCTGCTGATCGCCACGATGTGGGCGGCTCGGT 798
QY 1098 TGAATTCAAATCGCGATGCTTGGGCTTTGGCTTTATTTCTGTTTCAACCGGTGGTGG 1157
DB 799 GACCTTCGAGACGCGGATGCTTTCGCGGTGGGCTTCTGTCATCTGTTTCAACATCGCGG 858
QY 1158 TGTACCGGTATGCTGTCGCGCAAGCGGTCTGGACCGGTGATATCACCACCATATTA 1217
DB 859 CTTCTCGGACTGATGTGGGATCGCCCGGGGACTTCCAGTACACACACCTACTT 918
QY 1218 CGTGGTGGGACCTTCATTTATGATGTGCTGGGTGGATCTTTCGCTGCTTTCGCGG 1277
DB 919 CGTGGTGGGACCTTCATTTATGATGTGCTGGGTGGATCTTTCGCTGCTTTCGCGG 978
QY 1278 TATCTACTTTTACATGCGGATGCTTTCGCGCGGCTTTCGCGGATGGGCTGCAAGCT 1337
DB 979 GGTCTACTTGGCTGCGGAGTGGACCGGCGCATGTACGACGAGACCTTCGCGAGCT 1038
QY 1338 GCATCTTCGACCTTCTTCATCGGTGCGAAACGTCAGTTCCTTCCCGCAGCACTTCTCGG 1397
DB 1039 GCATCTTCGATGAGCTTCATCGGATGAACCTGGGCTTCTTCCCGATGACCTTCGCGG 1098
QY 1398 ACCTCAGGATGCGCGCGCTTACATCGACTATCGGAGCCCTTCGCGCTGCGGACAA 1457
DB 1099 CTTGCGCGGATGCGCGGCGGATCCCGGACTACAACTGCACTTCGCGACTTCAACAT 1158
QY 1458 AGTCTCTCTATGCTGGCTTCCTGGCTTCGCGCTTCGCTTCTTCTTCTTCTTCTTCTTCT 1517
DB 1159 GGTCTCTGATGCGCGCTTCATG--TTCCGACACCCAGCTGCTGTTCTGTTCTCAT 1215
QY 1518 TGTCTATACCTGGTGTGTCGCGCGCGGAGACCCCTTCGAAACCCCTGGGCGAAATTCG 1577
DB 1216 CGTCATCAAGTGCATCGCGCGGCGGCAAGCGGCGGCTTCGCAAGCCCTGGGAGCG--GCGC 1272
QY 1578 CGATACGCTGGAATGGAAGCTGCATCAGCGCTTCGCGGCGGCGGCGGCGGCGGCGGCGG 1637
DB 1273 CGAGGCGCTGGAGTGGAGCATCCCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1332
QY 1638 CAAG 1641
DB 1333 CGAG 1336

XX	29-JUL-2004	(first entry)	
XX			
XX			
DE			
DE			
KW			
KW			
XX			
OS			
PN			
PN			
PD			
XX			
XX	18-FEB-1999;	99US-00252991.	
XX			
PR	18-FEB-1998;	98US-0074788P.	
PR	27-JUL-1998;	98US-0094190P.	
XX			
PA	(GENO-)	GENOME THERAPEUTICS CORP.	
XX			
XX			
PI			
XX	Rubenfield MJ, Nolling J, Deloughery C, Bush D;		
DR			
DR	WPI; 2003-615309/58.		
DR	P-PSDB; ABO68501.		
XX			
PT	Novel isolated nucleic acid encoding <i>Pseudomonas aeruginosa</i> polypeptide,		
PT	useful as molecular targets for diagnostics, prophylaxis and treatment of		
PT	pathological conditions resulting from bacterial infection.		
XX			
PS	Disclosure; SEQ ID NO 676; 455pp; English.		
XX			
CC	The invention relates to <i>Pseudomonas aeruginosa</i> polypeptides and the		
CC	polynucleotides encoding them. The sequences are useful in diagnosis and		
CC	therapy of pathological conditions, as molecular targets for diagnostics,		
CC	prophylaxis and treatment of pathological conditions resulting from a		
CC	bacterial infection, for evaluating a compound, such as a polypeptide,		
CC	for the ability to bind a <i>P. aeruginosa</i> nucleic acid, as components of		
CC	effective antibacterial targets, as targets for antibacterial drugs,		
CC	including anti- <i>P. aeruginosa</i> drugs, as templates for recombinant		
CC	production of <i>P. aeruginosa</i> -derived peptides or polypeptides, as target		
CC	components for diagnosis and/or treatment of <i>P. aeruginosa</i> -caused		
CC	infection, and in detection of <i>P. aeruginosa</i> sequences or other sequences		
CC	of <i>Pseudomonas</i> species using biochip technology. Sequences ABD01397-		
CC	ABD17967 represent <i>P. aeruginosa</i> polynucleotides of the invention. Note:		
CC	The sequence data for this patent did not form part of the printed		
CC	specification but was obtained in electronic format from USPTO at		
CC	cc.seqdata.uspto.gov/sequence.html		
XX			
SQ	Sequence 1587 BP; 342 A; 458 C; 566 G; 221 T; 0 U; 0 Other;		
	Query Match	30.1%; Score 503.6; DB 11; Length 1587;	
	Best Local Similarity	63.5%; Pred. No. 1.4e-124;	
	Matches	84; Conservative 0; Mismatches 464; Indels 19; Gaps 4;	
QY	319	TTCCGGTGGTTTGTGTACTACTCTGATGCGCGGTGCAAAATCGGCGCTCCGGATATGG-CCTT	377
Db	1420	TTCTCGCGCCTGGCCAACTGGATGATCCCGTGTGATCGGCGCGCGGACATGGCCCT	1361
QY	378	CCCCGGTATGAACAACTGTGCTTTCGCTGCTTCATTCCGGTACCAGATGGCGCTGGC	437
Db	1360	GCOCGGATGAACAAATTCACCTTCGTGCTGCTCGCGCGGCCCTTCGGCCCTGCTGTGTCAG	1301
QY	438	TTCCGTGTTTCGACCGGCGGTGACGGTACAGCTGGGTTCGCGCGCTTGTTGGGTTCTGTA	497
Db	1300	CACCTGTTTCATGTCGCGCGCGGCCCAA-----CTTCGCTGGACCTTCTA	1253
QY	498	CCGCGCGCTGTGCAACCCGCGAAGCTGGCTATTCGATGGAACCTCGGATTTTCGGGTCA	557
Db	1252	TGCGCGCTGTGCAACACTTCGCCCCGCAACAGCGTGACCTTCCTTCATCTTCGCATCCA	1193
QY	558	CTTGTGCGGTGCTCTCGATCATGGGCGGATCAAAATGATCATCAGCACTTCCTTGAACAT	617

```
RESULT 5
ABD02134
ID ABD02134 standard; DNA; 1278 BP.
XX
AC ABD02134;
XX
DT 29-JUL-2004 (first entry)
XX
DE Pseudomonas aeruginosa polynucleotide #738.
XX
KW Bacterial infection; Gene; ds; Pseudomonas aeruginosa infection;
KW antibacterial.
XX
OS Pseudomonas aeruginosa.
XX
PN US6551795-B1.
XX
PD 22-APR-2003.
XX
PF 18-FEB-1999; 99US-00252391.
XX
PR 18-FEB-1998; 98US-0074788P.
PR 27-JUL-1998; 98US-0094190P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX
DR WPI; 2003-615309/58.
XX
DR P-PSDB; ABO68563.
XX
PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX
PS Disclosure; SEQ ID NO 738; 455bp; English.
XX
CC The invention relates to Pseudomonas aeruginosa polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a P. aeruginosa nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-P. aeruginosa drugs, as templates for recombinant
CC production of P. aeruginosa-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of P. aeruginosa-caused
CC infection, and in detection of P. aeruginosa sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences ABD01397-
CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html
XX
SQ Sequence 1278 BP; 182 A; 463 C; 361 G; 272 T; 0 U; 0 Other;
Query Match 29.2%; Score 489; DB 11; Length 1278;
Best Local Similarity 63.6%; Pred. No. 1.1e-120;
Matches 818; Conservative 0; Mismatches 450; Indels 19; Gaps 4;
QY 356 TCGGCGCTCGGATATGG-CTTCCCGCTATGAACAACCTTCGTTCTGGCTGTTCATT 414
DB 1 TCGGCGCGCGGACATGGCCCTTCGCGGATGAACAATTTTCAGCTTCGCTGTGCGG 60
QY 415 GCGGCTACCGGATGGCGTGGCTTCGCTTCGACCGCGGCGGTGACGGTCAGCTGGGT 474
DB 61 GCGGCTTCGGCTGCTGTGTACGACCCCTGTTCATGCCCGGCGGCGCCAA----- 113
QY 475 TCGGGCGTGGTGGTTCGTATCCCGCGCTGTTCGACCCGCGAAGCTGGCTATTTCGATG 534
DB 114 -----CTTCGGCTGGACCTTCTATGGCGCGCTGTTCGACCACTTCGCGCGCACAGCGTG 168
```

Qy	1615	GCCACACACCTTCGAAACGCTGCCCAAG	1641
Db	1243	TACCACACCTTCAGCACCCGCCGAG	1269
RESULT 6			
ACC45136	ACC45136 standard; DNA; 16568 BP.		
XX	ACC45136;		
XX	XX		
DT	11-JUN-2003	(first entry)	
XX	XX		
DE	Human mitochondrial DNA sequence SEQ ID NO:1.		
XX	XX		
KW	Human; mitochondrial DNA; mtDNA; allele; diagnosis; haplotype;		
KW	criminal investigation; forensic analysis; energetic imbalance;		
KW	energy metabolism-related physiological condition; metabolic disease;		
KW	abnormal energy metabolism; abnormal temperature regulation; diabetes;		
KW	abnormal oxidative phosphorylation; abnormal electron transport; obesity;		
KW	hypertension; cardiovascular disease; gene; ds.		
XX	XX		
OS	Homo sapiens.		
XX	XX		
FT	Key	Location/Qualifiers	
FT	misc_feature	1..16568	
FT		/tag= a	
FT		/note= "Nucleotide alleles are given using the IUPAC	
FT		ambiguity code bases e.g. y for T or C"	
XX	XX		
FN	WO2003018775-A2.		
XX	XX		
PD	06-MAR-2003.		
XX	XX		
DF	30-AUG-2002; 2002WO-US028471.		
XX	XX		
PR	30-AUG-2001; 2001US-0316333P.		
PR	31-AUG-2001; 2001CA-02356536.		
PR	13-MAY-2002; 2002US-0380546P.		
XX	XX		
PA	(UYEM-) UNIV EMORY.		
XX	XX		
FI	Wallace DC, Hosseini S, Mishmar D;		
XX	XX		
DR	WPI; 2003-290064/28.		
XX	XX		
PT	Diagnosing a haplogroup from a human, useful in criminal investigations		
PT	and forensic analyses by providing a sample having mitochondrial nucleic		
PT	acid, and identifying the presence of nucleotide allele/s diagnostic of a		
PT	haplogroup.		
XX	XX		
PS	Claim 23; Page 117-121; 126pp; English.		
XX	XX		
CC	The present invention describes a method for diagnosing a haplogroup from		
CC	a human comprising providing a sample comprising mitochondrial DNA		
CC	(mtDNA) from the human, and identifying in the sample, the presence of at		
CC	least one nucleotide allele diagnostic of a haplogroup. The method of		
CC	diagnosing haplogroups of a human is useful in criminal investigations		
CC	and forensic analysis. Arrays from the present invention are useful for		
CC	determining the presence or absence of nucleotide alleles. The alleles		
CC	are useful for diagnosing predisposition to an energy metabolism-related		
CC	physiological conditions such as energetic imbalance, metabolic disease,		
CC	abnormal energy metabolism, abnormal temperature regulation, abnormal		
CC	oxidative phosphorylation, abnormal electron transport, obesity,		
CC	diabetes, hypertension, or cardiovascular disease. The present sequence		
CC	represents a human mtDNA sequence which shows alleles in the sequence		
CC	with IUPAC ambiguity code bases e.g. y for T or C		
XX	XX		
SQ	Sequence 16568 BP; 4859 A; 4950 C; 1991 G; 3829 T; 0 U; 939 Other;		
Query Match 24.4%; Score 408.4; DB 8; Length 16568;			
Best Local Similarity 56.8%; Pred. No. 1.2e-98;			
Matches 726; Conservative 32; Mismatches 508; Indels 12; Gaps			

QY 1321 GAATGGGCTGCAAGTGCACCTTCTGACCTTCTTATCGGTGCGAAGTCACGTTCTTC 1380
DB 7121 CAACACCTACGCAAAATCCATTTTCCTATCATATTCATCGGCGTAAATCTAACVTTCTTC 7180
QY 1381 CCGCAGACATCTCTGGGACGTCAGGATGCGCGCGTTACATCGCACTATCCGAGCC 1440
DB 7181 CCACACATTTCTGCGCTTCCTCGGAATGCCCGAGCTACTCGACTACCCCGATGCA 7240
QY 1441 TTCGCGCTGGAAACAAGTCTGCTCTATGTCGCTTCTGCGCTTCTGCTCGCTCGTTCCTG 1500
DB 7241 TACACACATGAAYRTCTATCATCTGTRGGVTCATTTCATTCTCTAACAGCAGTAATA 7300
QY 1501 TTCTTCATCTGATCTTT 1518
DB 7301 TTAATAATTTTCATGATY 7318

RESULT 7

ACA37676
ID ACA37676 standard; DNA; 1629 BP.

XX ACA37676;

AC ACA37676;

DT 19-JUN-2003 (first entry)

XX Prokaryotic essential gene #19333.

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

XX drug design; gene.

XX Mycobacterium avium.

OS WO200277183-A2.

PN 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00815242.

XX 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

XX WPI; 2003-029926/02.

DR P-PSDB; ABU33806.

XX New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

XX Claim 14; SEQ ID NO 25546; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies

CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: the sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX
SQ Sequence 1629 BP; 218 A; 560 C; 505 G; 346 T; 0 U; 0 Other;

Query Match 24.3%; Score 406.2; DB 8; Length 1629;

Best Local Similarity 59.0%; Pred. No. 2e-98;

Matches 800; Conservative 0; Mismatches 533; Indels 24; Gaps 5;

QY 288 GATGATGTTCTTTGTGGGTATCCCGCATTTGTCGGTGGTTTGGTAACATATCTGATGCC 347

DB 150 GATCATGCTGCTGTACGCCACCCCGGTGTTTCGGTTTCGCCAACCTGGTGTGCC 209

QY 348 GCTCAAAATCGCGCTCGGATATGGCTTCCCGGTATGAACAACTGCTGTTCTGGCT 407

DB 210 GCTGAGATCGCGCGCCGACGTGGCTTCCCGCGCTCAACGCCCTTCTGTTCTGGCT 269

QY 408 GTTCATTGCCGCTACCGCATGCGCGGTTCGCTTCGCACCGCGCGGTGACGTCA 467

DB 270 GTTCTGTTCCGCGGCTGATCGCGGCTCGGCTTCATGTCGCGCGCGCGCGCCGA 329

QY 468 GCTGGGTTCCGGCGTTGGTTGGTTCTGTACCCCGCGCTGTCGACCCCGCAAGTGGCTA 527

DB 330 CTTCGGCTGGACCGCTACACCCCGCTGTGACCGCGTGCACCTGCGCGCGCGCGG--- 386

QY 528 TTGATGACCTCGGATTTTCGGGTTCACTTGGGTGCTCTCTCGATCATGCGGC 587

DB 387 ---GGGAGACCTGAGATCACCGGCTGATGCTGCGCGGCTCTGGGACCATCTCTGGGTGC 443

QY 588 GATCAACATGATCAGACCTTCTTGAACATGCGCGCGCGCGCGCATGACGTCACAAAGT 647

DB 444 GGTCAACATGATCACACCGCTGCTCATGCGCGCGCGCGCATGACGATGTTCCGAT 503

QY 648 GCGGTTGTTCTGTTGTCGATTTTATCAGGCTTGGCTGATCTCTGCTGGCGCTGCCGT 707

DB 504 GCGGATCTTCACTTGAACATCTCTGGTGCATCTGATCTGATCTGATCTGATCTGCGAT 563

QY 708 TCTGGCTGGTGCATCACCATCTGCTGACCGCGCTTGGCATCATCAGCCATCTGCTGACCTT 767

DB 564 CTTGACCGCGCGCTGTTGGGCTGGCGCGCGCGCGCATCTGGCGCGCATGTTACGA 623

QY 768 TCTGCTGGCGCGGTGACCGGATCTGTACCAACACATCTCTGTTGTTCTTGGGACCC 827

DB 624 CGCGCGCAACCGCGGAGTTC---TGTTGGGCGACCATCTGTTGGTCTCTGCGCATCC 680

QY 828 GGAAGTGTACATCATCTGTCGCGGCTTGGCATCATCAGCCATCTGCTGTCGACCTT 887

DB 581 CGAGGTGTACATCATCTGTTGCGGTTCCGTTCTTCGGGATCATCAGAGATCATCCGGTGT 740

QY 888 CTCGAAAGCCGCTTTCGGTTTACCTGCGGATGTTCTATGCAATGTTGGCAATCGGTGT 947

DB 741 CGCGCGCAACCGCGGAGTTC---TGTTGGGCGACCATCTGTTGGTCTCTGCGCATCC 800

QY 948 TCTGGGCTTGTCTGTCGGCGCACCATGTACCGTTGGTATGTCGTCGACCCAGCA 1007

DB 801 GCTTCGGTTCGGGTGCGGCGCACCATGTGTCGACCGGAGCGTCTCTGTCGCGTT 860

QY 1008 ATCTACTTTCATGTCGCGCACCATGTGTCGCGGTGCGCGCATTAAGATCTTCTC 1067


```
Db 861 CTTTTCGTTCTG-----ACGTACTGATCGCGGTGCGGACCGGATCAAGTTCTTCAA 914
Qy 1068 GTGATGCGCCAGATGTGGGGCGGTGCGTTGAGTTCAAAATCGCGATGCTCTGGGCCCTT 1127
Db 915 CTGGATGCGCAGATGTGGAAGGCGAGTTGACATTTGAGACGCGATGCTGTTCTGGGT 974
Qy 1128 TGGCTTTATGTTCTGTTACCGTGGTGGTGTGACCGGTATCGTGGCCCAAGCGGG 1187
Db 975 CGGCTTCTGCTCACTTCTGCTGGTGGTGTGACCGCGGTGATGCTGGCCAGCCGCC 1034
Qy 1188 TGTGACCGTGCATATCAGACACCTATTACGTGTGGCGCACTTCCCATTTATGATGTC 1247
Db 1035 GTGGACTTCCAGTCACCGACACCTATTTCGTGGTGGCGCACTTCCACTACGTGCTGT 1094
Qy 1248 GTGGGTGGATCTTGGATCTTGGCGGTATCTACTTTACATGCCGAAGTCTCGGG 1307
Db 1095 CGGCACCAATCGTGTTCGCCACCTTCGCGGGGTCTACTTCTGGTTCGCCGAAGATACCGG 1154
Qy 1308 CGCGCTTTCGCGAATGGGCTGCAAGGTGCACTTCTGGACCTTCTTCATCGGTGCGAA 1367
Db 1155 CGGCTGCTGACGAGCGGCTGGGCAAGTGTCACTTCTGTTGACGTTTCATCGGTTCCA 1214
Qy 1368 GGTACGTTTCTCCGACGACTTCTCTGGACGTCAGGGTATGCGCGCGGTTACATCGA 1427
Db 1215 CACCACCTTCTGTGTCAGCACTGCTGGCGGATTTGGCGCATGCGCGCGCTACGCCGA 1274
Qy 1428 CTATC-----CGAAGCCTTCGCGTGTGGAAACAAAGTCTGCTCTATGTTGCGTTCCT 1481
Db 1275 CTACTGCCAGCGAGCGGTTTCAGCCCTTACAGCTGCGCTTCCAGCTGCGGGCGTTTAT 1334
Qy 1482 GGCCTTCGCTCTGTTCTCTTCTCATCGTATCTTTGTCTATACGCTGTTGTGTCGCG 1541
Db 1335 CTGGGCGGCTCGATGTTCCCGTTCGTCTGGAACGTTCTCAAGAGTGGCGCTACGGCGA 1394
Qy 1542 CGCGAGACCGTCCGAACCGTGGGCGAATTCGCCGATAGCTGGATGACGCTGCC 1601
Db 1395 GGTGCTGTCAGCGTGCAGACCCGCTGGG-----TTACGGCAACTCGCTGGAGTGGCGACCA 1451
Qy 1602 ATCACCCTCGCTCGGCGCCACAGCTTCGAACCGTGGCC 1638
Db 1452 CTCCCGCGCGCGCGGCGACAACTTCACCGAGCTGCC 1488

RESULT 8
AAS75523
ID AAS75523 standard; cDNA; 1555 BP.
AC AAS75523;
XX
DT
DE 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #11327.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
FN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR P-PSDB; ABG11336.
DR
```

```
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 11327; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activities. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1555 BP; 416 A; 469 C; 253 G; 417 T; 0 U; 0 Other;
Query Match 24.2%; Score 405; DB 5; Length 1555;
Best Local Similarity 57.9%; Pred. No. 4e-98;
Matches 740; Conservative 0; Mismatches 535; Indels 4; Gaps 1;
Qy 244 AACGCACTGTGGACGTCATGTTACCTACCATGTTTCTGATGATGTTCTTGTG 303
Db 151 AACGACCACTATACAAACGTTATGTCACAGCCCATGTCATTGTAATAATCTTCTCATA 210
Qy 304 GGTATCCCGCATTTGTTGGTGGTTTGGTAACATATCTGATGCGCTGCAAAATCGGCGCT 363
Db 211 GTAATACCATCATATAATCGGAGGCTTTGGCACTAGTTCCTCTAATAATCGGTGCC 270
Qy 364 CCGGATATGCGCTTCCGCGATGAAACAACTGTGTTGTTGGTGTTCATTCGCGGTACC 423
Db 271 CCGGATATGCGCTTCCCGCGATAAACAACTAGCTTCTGACTCTTACCTCCCTCTCTC 330
Qy 424 CGGATGGCGTGGCTTCGCTGTTCCGACCGGGGTGACGTCAGCTGGGTTCGGGCGTT 483
Db 331 CTACTCTGCTCGCATGTCGTATAGTGAAGCGCGGAGCAGGAACAAGTTTGAACAAGT 390
Qy 484 GTTTGGGTTCTGTACCCGCGCTGTGACCCGCGAAGCTGGCTATTCGATGGACCTCGCG 543
Db 391 CGTACCTCCCTTTAGCAGGAACTACTCCACCTTGAAGAGCTCCGGTAGACCTAAC 450
Qy 544 ATTTTCGGGTTCACTTGTGGGTGCTCTCGATCATGCGCGCGATCAACATGATCAG 603
Db 451 ATCTCTCCTTACACCTAGCAGGTGTCTCTCTATCTTAGGGCCATCAATTTTCATACA 510
Qy 604 ACCTTCTTGAACATCGCGCGCCCGCATGACGTCGACAAAGTGCCTGTTTCTCGTGG 663
Db 511 ACAATATCAATATAAAACCCCTGCGCATAAACCAATACCAACGCGCTCTCTGCTGA 570
Qy 664 TCGATCTTTATCAGCGCTTGGCTGATCTGCTGGCGCTGCGGTTCTGGGTGGTGCATC 723
Db 571 TCCGTCTCTAATCACAGCAGTCTACTTCTCTCTATCTCTCCAGTCTCTAGTGTGGCATC 630
Qy 724 ACCATGCTGTGACCGCGGTAACTTCGGCACGACCTTCTTCAATCTGTGGGCGGCT 783
Db 631 ACTATACTACTAACAGACCGCAACTCAACACCACTTCTTCGACCCCGCGGAGGAGA 690
Qy 784 GACCCGATTTCTGTACCAACACATCCTGTGGTCTTTTGGGACACCGGAAGTGTATCATC 843
```



```
Db 6289 -----TACCTCCCTTAGCAGGAACTACTCCACCCCTGGAGCCTCCGTAGACCTAAC 6342
Qy 544 ATTTTCGGGGTTCACTTCGGGTGCTCTTCGATCATGGCGGATCAACATCATCAAG 603
Db 6343 ATCTTCTCTTACACCTAGAGGTGTCTCTCTAATCTTAGGGCCATCAACTTCATCACA 6402
Qy 604 ACCTTCTTGAACATCGCGGCCCGGATGACGCTGCACAAAGTCCGTTGTCTCGTGG 663
Db 6403 ACAATTATCAATAAACCCTCGCCATAACCAACCAACCAACCAACCAACCAACCAAC 6462
Qy 664 TCGATCTTTATCAGCGCTTGCATCTCTGCTGGCGGTCTCTGCTGCTGCTGCTGCTGCT 723
Db 6463 TCGCTCTTATCAGCAGCTCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6522
Qy 724 ACCATGCTGTGACCGACCGTAACTTCGGCACGACCTTCTTCAATCTGCTGGCGCGGT 783
Db 6523 ACTATCTACTAACAGACCGCACTCAACACACCTTCTTCGACCCCGCGGAGGGA 6582
Qy 784 GACCGGATCTGTACACACATCTCTGTGTTCTTTGGCACCGGAGTGTACATCATC 843
Db 6583 GACCCCATCTTATACCAACCTATCTGATTTTTCGGTCAACCTGAAGTTTATATCTT 6642
Qy 844 ATCTGCGCGCTTGTGGCATCATCAGCCATGTCGTGTGACCTTCTC- --GAAAAAGCCG 900
Db 6643 ATCTGACCGGCTTCGGAATAATCTCCATATTTGTAATCTTACTCTCGGAAAGAA 6702
Qy 901 GTCTTCGTTACCTGCGCATGTGCTATGCAATGTGGCAATCGGTGTTTGGGCTTTGTC 960
Db 6703 CCATTTGGATACATAGTATGTTGCTGAGCTATGATATCAATTTGCTTCTTAGGTTTATC 6762
Qy 961 GTCTGGCGCACCATGTACACCGTTGTTGATGTCGCTGACCCAGCAATCTTACTTCATG 1020
Db 6763 GTGTGACACACATATATTACAGTAGAATAGACGTAGACACACAGCATATTTTACC 6822
Qy 1021 CTGGCCACCATGTTGATGCGGGTGCACCGGATTAAGATCTTCTGTGATGATCGGCACG 1080
Db 6823 TCCGCTACCATATCATCGCTATCCCGCGGGTCAAGTATTTAGCTGACTCGCCACA 6882
Qy 1081 ATGTGGCGCGCTGCTGTAGTTCAATCGCGATGCTCGGCTTGGCTTTGTTGTTTC 1140
Db 6883 CTCACGAGCAATATGAATGATCTGTGCTGAGTCTCTGAGCCCTTAGGATTCATCTT 6942
Qy 1141 CTGTTCACCGTGGTGTGACCGGTATCGTGTGCGCCCAACGCGGTCTGACCGTGA 1200
Db 6943 CTTTTCACCGTAGTGGCTGACTGGCATTTGTTAGCAAACTCATCACTAGACATCGTA 7002
Qy 1201 TATCAGCACACCTATTACGTTGGCGCATCTTCATATGTGATGTGCTGGGTGGATC 1260
Db 7003 CTACAGCACACGCTACTAGTTGAGTCACTTCCATGTTCTCTATCAATAGAGCTGA 7062
Qy 1261 TTTGCGATCTTCGCGGTATCTACTTTTACATCGCGAAGTTCTCGGCGCGCTTTCGCG 1320
Db 7063 TTTGCCATCATAGAGGCTTCATCTAGTTTCCCTATTCTCAGGCTACACCTTAGAC 7122
Qy 1321 GAATGGCTGCAAGCTGCACTTCTGCACTTCTTCATCGTGGCAAGCTACGTTCTTC 1380
Db 7123 CAACCTTACGCCAAAATFCAATTTCACTATCATATTCATCGGCGTAAATCTAATCTTCT 7182
Qy 1381 CCGCAGCATTCTCGGGAGCTCAGGCTATGCGCGCGCTTACATCGACTATCCGAGCC 1440
Db 7183 CCACACATTTCTCGCCTATCCGNAATGCCCGGAGCTTCTCGGACTACCCCGATGA 7242
Qy 1441 TTGCGCTGTGGACAAAGTCTCTGCTCTATGTGCTTCTGCGCTTCTGCGCTGTTCTG 1500
Db 7243 TACACCACATGAACATCTCTATCATCTGTAGGCTCATCTTCTCTTCAACAGCAGTAATA 7302
Qy 1501 TTCTTCATCGTATCTTT 1518
Db 7303 TTAATAATTTTCATGATT 7320
```

```
ADD43240
ID ADD43240 standard; DNA; 16568 BP.
XX
AC ADD43240;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human mitochondrial DNA (mtDNA) SEQ ID NO:414.
XX
KW mitochondrial haplogroup; mitochondrial DNA; mtDNA;
KW single nucleotide polymorphism; SNP; genetic relationship; antidiabetic;
KW neoplastic; neuroprotective; cytosolic; gene therapy; genealogy;
KW forensic; Alzheimer's disease; cancer; type 2 diabetes mellitus; human;
KW ds.
XX
OS Homo sapiens.
XX
PN WO2003046225-A1.
XX
XX 05-JUN-2003.
XX
PF 25-NOV-2002; 2002WO-US038276.
XX
PR 26-NOV-2001; 2001US-0333622P.
PR 28-MAR-2002; 2002US-036913P.
PR 01-APR-2002; 2002US-0369539P.
XX
XX (MITO-) MITOKOR.
XX
XX Herrnsstadt C;
XX
XX WPI; 2003-505214/47.
XX
PT Determining single nucleotide polymorphisms in mtDNA or homoplasmic mtDNA
PT mutations, useful for diagnosing and treating diseases, such as
PT Alzheimer's disease, cancer and type 2 diabetes mellitus.
XX
PS Example 2; SEQ ID NO 414; 193pp; English.
XX
CC The present invention describes a method (M1) for determining the
CC mitochondrial haplogroup of a subject, comprising determining in a
CC biological sample with mitochondrial DNA (mtDNA) from a subject, the
CC presence or absence of at least one mitochondrial single nucleotide
CC polymorphism (SNP) that is associated with a mitochondrial haplogroup.
CC Also described: (1) determining a genetic relationship between two
CC subjects; (2) determining a genetic relationship between an unknown
CC source or biological subject from which an unidentified sample is
CC obtained, and a known source or biological subject from an identified
CC sample is obtained; and (3) determining the presence of or the risk of
CC having a disease associated with a mtDNA SNP. Mitochondrial DNA can have
CC antidiabetic, neoplastic, neuroprotective and cytosolic activities, and
CC can be used in gene therapy. M1 and compositions of the present invention
CC are useful for detecting the presence or risk of diseases, treating such
CC diseases, determining the haplogroup of an individual, and establishing
CC genetic relationships between individuals for genealogical and forensic
CC purposes. The diseases include Alzheimer's disease, cancer and type 2
CC diabetes mellitus. The present sequence represents a full length human
CC mtDNA sequence, which is used in the exemplification of the present
CC invention.
XX
SQ Sequence 16568 BP; 5116 A; 5186 C; 2176 G; 4089 T; 0 U; 1 Other;
```

Query Match 23.8%; Score 398.8; DB 10; Length 16568;
Best Local Similarity 58.2%; Pred. No. 4.6e-96;
Matches 744; Conservative 0; Mismatches 522; Indels 12; Gaps 2;

```
Qy 244 AACGACACCTGTGGAACGTCATGTTACCTACCATGTTTCTGATGATGTTCTTGTG 303
Db 6050 AACGACACATCTCAACGTTATCGTCAGCCCATGATTTGTAATAATCTTCTTCATA 6109
Qy 304 GGATATCCCGCATTTGTTGCGTGGTTTGGTAACTATCTGATGCGGTGCAAAATCGGCGCT 363
Db 6110 GTAATACCCATCATATCGGAGGCTTTGGCACTGACTAATCCCTTAATAATCGGTGCC 6169
```

QY	364	CGGATATGGCCTTCCCGGTATGAACAACAGTCGTGTTCTTGCTGTATTCAITGCCGGPACC	423
Db	6170	CCGATATGGCGTTTCCCAGATAAACAACAATAGCTTCTGACTCTTACCTCCCTCTCTC	6229
QY	424	SGCATGGSCGTGGCTGCCTGTTTCGCACCGGGCGGTGACGGTCACTGGGTTCGGGGGTT	483
Db	6230	CTACTCTCTGCTCGCATCTGCTATAGTAGGAGGCCGAGGACAGGTGTGAACAGTC---	6286
QY	484	GTTGGGTTCTGTACCCCGCGCTGTGCAACCGGAAGCTGGCTATTGATGGACCTCGG	543
Db	6287	-----TACCCTCCCTTAGCAGGAATACTTCCACCTCGAGGCTCGTAGACCTTAACC	6340
QY	544	ATTTTCGGGGTTCACTTGTTCGGGTGGCTCTCTCGATCATGGCGGATCAACATGATCACG	603
Db	6341	ATCTCTCTTACACCTAGCAGGTGCTCTCTATCTTTAGGGGCCATCAATTTCATCACA	6400
QY	604	ACCTCTTTGAACATGCGCGCCCGGCATGACGCTGCACAAAGTCGCTGTCTCGTGG	663
Db	6401	ACAAATATCAATAPAAAACCCCTGGCATAAACCAATACCAAACGCCCTCTTCTGTCTGA	6460
QY	664	TGATCTTTATCAGGCTTGGCTGATCTCTGCTGGGCTGCCGGTTCGTGTGTGCAATC	723
Db	6461	TCCGTCTTAATCACAGAGTCTTACTTCTCTTCTCTCTCTCTCCAGTCTAGCTGTGGCATC	6520
QY	724	ACCATGCTGTCACCGACCGTAATCTTGGCAGACCTCTTCAATCTGCTGGCGGGT	783
Db	6521	ACTATACTATAAGAACCCCACTCAACACCACTTCTTGACCCCGCCGAGGAGA	6580
QY	784	GACCGATTCGTATCAACAACATCTCTGTGGTCTTTTGGGACCCGGAAGTGTACATCATC	843
Db	6581	GACCCATCTTATACCAACACCTATTCTGATTTTTCGGTCACTGAGTTTATATTTCTT	6640
QY	844	ATTTGCGCGGCTTTGGCATCATCAGCCATGCTGTGTGACCTTCTC---GAAAAAGCCG	900
Db	6641	ATCCTACAGGCTTCGGAATAATCTCCCATATGTGTAATTACTACTCOGGAANAAGAA	6700
QY	901	GTCTTCGTTACCTCCGANGGTCTATGCAATGGTGGCAATCGGTGTTCTGGGCTTTGTC	960
Db	6701	CCAATTGGATACATAGTATGGTCTGAGTATGATATCAATTTGGTCTCTAGGGTTATC	6760
QY	961	GTCTGGGCGCACCATATGACCCGTGGTATGCTGCTGACCCAGCAATCTTACTTCATG	1020
Db	6761	GTGTGAGCACACCATATATTACATAGGAATAGACGTAGACACACAGACATATTTCAAC	6820
QY	1021	CTGGCCACCATGGTATCGCGGTGCCGACCGGCATTAGATCTTCTGTGGATCGCCACG	1080
Db	6821	TCCGCTACCATATCATCTGATCTCCCAACGGGCTCAAGTATTTAGCTGACTCGCCACA	6880
QY	1081	ATGTGGGCGGCTCGGTTGAGTTCAAAATGCCAGTCTCTGGGCTTTGGCTTTATGTTTC	1140
Db	6881	CTCCACGGAACAAATATGAATGATCTGTGCAGTGTCTCGAGCCCTAGATTCATCTTT	6940
QY	1141	CTGTTTACCGGTGGTGTGACCGGTATCTGCTGGCCCCAAGGGGTCTGGACCGTGCA	1200
Db	6941	CTTTTACCGTAGTGGGCTGACTGGCATTTGATTAGCAAACTCATCACTAGACATCGTA	7000
QY	1201	TATCAGACACCTATACGTGTGGCGCATTTCCATTTATGTATGTGCTGGGTGGGATC	1260
Db	7001	CTACACGACAGTACTAGTTGTAGTCTACTTCCACTATGTCTTATCAATAGSAGCTGTA	7060
QY	1261	TTTTCGATCTTCGCGGCTATCTACTTTTACATGCCGAAGTCTCGGGCGGGCTTTCCCG	1320
Db	7061	TTTGCCATCATAGGAGGCTCAITTCATCTGATTTCCCTATTCTCAGGCTACACCTAGAC	7120
QY	1321	GAATGGGCTCAAGAGCTGCATCTTCGACCTTCTTCATCGGTGCGAAGCTCAGTCTTTC	1380
Db	7121	CAAACTACGCCAAAATCCAAATTTCACTATCATATTATCATCGGCGTAAATCTTAACCTT	7180
QY	1381	CCGACGACTTCTCGGACGCTCAGGATATGCCGCGCGTTATCATCGATATCCCGAAGCC	1440
Db	7181	CAACAACATCTTTCGGCTTATCGGAATGCCGACGTTACTCGGACTACCCGATGCA	7240

Qy	1441	TTGGCGCTGTGGAAACAAAGTCTGGTCCTATGTTGGTGCGTCTCCTGGCCCTTCGCCTCGTTCCTG	1500
Db	7241	TACACCACATGAACAATCATCTATCATCTGTAGGCTCATTCATTCTCTAAACAGCAGTAATA	7300
Qy	1501	TTCTTTCATCGTGATCTTT	1518
Db	7301	TTAATAATTTCATGATT	7318
 RESULT 12 ADD43266 ID ADD43266 standard; DNA; 16569 BP.			
XX	XX	XX	XX
XX	XX	AC	AC
XX	XX	ADD43266;	
DT	DT	15-JAN-2004	(first entry)
XX	XX	Human mitochondrial DNA (mtDNA) SEQ ID NO:440.	
DE	XX	mitochondrial haplogroup; mitochondrial DNA; mtDNA;	
KW	KW	single nucleotide polymorphism; SNP; genetic relationship; antidiabetic;	
KW	KW	nootropic; neuroprotective; cytostatic; gene therapy; genealogy;	
KW	KW	forensic; Alzheimer's disease; cancer; type 2 diabetes mellitus; human;	
XX	XX	ds.	
XX	XX	Homo sapiens.	
OS	XX		
XX	XX	WO2003046225-A1.	
PN	XX		
PD	XX	05-JUN-2003.	
PF	XX		
XX	XX	25-NOV-2002; 2002WO-US038276.	
XX	XX		
PR	XX	26-NOV-2001; 2001US-0333622P.	
PR	XX	28-MAR-2002; 2002US-036931P.	
PR	XX	01-APR-2002; 2002US-0369539P.	
XX	XX	(MITO-) MITOKOR.	
PA	XX		
XX	XX	Herrnstadt C;	
Pf	XX		
XX	XX	WPI; 2003-505214/47.	
DR	XX		
XX	XX	Determining single nucleotide polymorphisms in mtDNA or homoplasmic mtDNA	
PT	XX	mutations, useful for diagnosing and treating diseases, such as	
PT	XX	Alzheimer's disease, cancer and type 2 diabetes mellitus.	
PS	XX	Example 2; SEQ ID NO 440; 193pp; English.	

The present invention describes a method (M1) for determining the mitochondrial haplogroup of a subject, comprising determining in a biological sample with mitochondrial DNA (mtDNA) from a subject, the presence or absence of at least one mitochondrial single nucleotide polymorphism (SNP) that is associated with a mitochondrial haplogroup.

Also described: (1) determining a genetic relationship between two subjects; (2) determining a genetic relationship between an unknown source or biological subject from which an unidentified sample is obtained, and a known source or biological subject from an identified sample is obtained; and (3) determining the presence of or the risk of having a disease associated with a mtDNA SNP. Mitochondrial DNA can have a disease associated with it. M1 and compositions of the present invention are useful for gene therapy. M1 and compositions of the present invention are useful for detecting the presence or risk of diseases, treating such diseases, determining the haplogroup of an individual, and establishing genetic relationships between individuals for genealogical and forensic purposes. The diseases include Alzheimer's disease, cancer and type 2 diabetes mellitus. The present sequence represents a full length human mtDNA sequence, which is used in the exemplification of the present invention.

XX
SQ Sequence 16569 BP; 5115 A; 5174 C; 2177 G; 4102 T; 0 U; 1 Other;

Query Match 23.8%; Score 398.8; DB 10; Length 16569;

Best Local Similarity 58.2%; Pred. No. 4.6e-96;
Matches 744; Conservative 0; Mismatches 522; Indels 12; Gaps 2;

QY 244 AACGACACCTGTGGAAGCTGATGGTTACCTACCAATGTAATTCGATGATGTTCTTTGTG 303
Db 6051 AACGACACCTGTGGAAGCTGATGGTTACCTACCAATGTAATTCGATGATGTTCTTTGTG 6110
QY 304 GGTATCCCGCATGTTTCGGTGGTTTTCGTAACATATCTGATGCGCTGCAATCGGCGCT 363
Db 6111 GTATATCCCATCAATGCGGAGGCTTTGGCACTGATGTTCCCTTAATAATCGGTGCC 6170
QY 364 CGGATATGGCTTCCCGCATGTAACCACTGCTGTTCTGGCTGTTTCATTCGCGGTACC 423
Db 6171 CCGGATATGGCTTCCCGCATGTAACCACTGCTGTTCTGGCTGTTTCATTCGCGGTACC 6230
QY 424 CGGATGGCGGTGCTGCTGTTTCGACCGCGGCTGACGTCAGCTGGGTTGGGCGTT 483
Db 6231 CTATCTCGCTCGCATCTGCTATAGTGAGCGCGGAGCAGGATGAAAGTC--- 6287
QY 484 GGTGGGTTCTGTACCGCGCTGTGACCGCGAAGCTGGCTATTCGATGGACCTCGCG 543
Db 6288 -----TACCCCTCTTAGCAGGAACTACTCCACCCCTGGAGCTTCGTAGACCTAAC 6341
QY 544 ATTTTCGGGTTCACTTGTGGGTCCTCTCGATCATGGGCGGATCAACATGATCAAG 603
Db 6342 ATCTCTCTTACACCTAGCAGGTTCTCTCTATCTTAGGGCCATCAATTTTCATCACA 6401
QY 604 ACCTTCTTGAACATGCGCGCGCCCGCATGACGCTGCACAAAGTGCGGTTGTTCTCGTG 663
Db 6402 ACAATTAATCAATAAACCCTGTCATACCCATACCAAGCCCTCTTCTGCTGA 6461
QY 664 TCGATCTTATCAGCGTTGGCTGATCCTCTGCGCGCTGCGGTTCTGGCTGTCGAAATC 723
Db 6462 TCCGTCCTAATACAGCAGTCTACTTCTCTATCTCTCCAGTCTAGTGTGGCATC 6521
QY 724 ACCATGCTGTGACCGACCGTAACTTCGGCACGACCTTCTCAATCTGCTGCGCGGCT 783
Db 6522 ACTATATCTACAGACCGCACTCAACACACCTTCTTCACCCCGCGGAGGAGA 6581
QY 784 GACCGATCTGTACCAACACATCTGTGTTCTTTGGCACCCGGAAGTGATCATC 843
Db 6582 GACCCCATCTATACCAACACCTATCTGATTTTCGCTCACCTGAAGTTTATATCTT 6641
QY 844 ATCTGCGCGCTTTGSCATCATCAGCATGCTGCTGCTGACCTTCTC---GHAAGACCG 900
Db 6642 ATCTGACCGCTTGGAAATCTCCCATTTGTAACTTACTTACCCGGAAGAAAGAA 6701
QY 901 GTCTTGGTTTACCTGCGGATGCTATGCAATGGTGGCAATCGGTTCTTGGGCTTTGTC 960
Db 6702 CCATTTGGATACATAGTATGGTCTGAGTATGATATCAATTTGGCTTCTAGGGTTATC 6761
QY 961 GTCTGCGCGCACCATGTACACCGTTGATGCTGCTGACCCAGCAATCTTACTTATG 1020
Db 6762 GTGTGAGCACCATATATTTACAGTAGAATAGACGTAGACACAGCATATTTTACC 6821
QY 1021 CTGCGCACCATGTTGATGCGGTCGCGGATTAAGATCTTCTCGGTTGATTCGCCAG 1080
Db 6822 TCCGCTACCATATCATCTGCTATCCCGACCGGCTCNAAGTATTTAGTGTACTGCCACA 6881
QY 1081 ATGTGGGGGCTCGGTTGAGTTCAATCGCGATGCTCTGGGCTTTGGCTTTATGTTTC 1140
Db 6882 CTCACGGAAGCAATATGAATGATGCTGCTGAGTGTCTGAGCCCTAGGATTTATCTT 6941
QY 1141 CTGTTACCGCTGGTGTGACCGGTATCTGCTGCGCCCAAGCGGCTGTGACCGTGA 1200
Db 6942 CTTTTCACCGTAGTGGCTGACTGCGATGTTATAGCAAACTCATCTAGACATCGTA 7001
QY 1201 TATCAGCACCTATTTACGTGGGCGCACTTCCATTTATGATGCTGCGGTGGGATC 1260
Db 7002 CTACACGACACGCTACTACGTTGTAGTCTACTTCCACTATGCTCTATCAATAGGACCTGA 7061
QY 1261 TTTGCGATCTTCCGCGGTATCTACTTTTACATGCGGAGTCTCGGCGCGCTTTCCCG 1320

Db 7062 TTTGCCATCATAGGAGGCTTCAATTCAGTATTTCCCTATTTCTCAGGCTACACCTAGAC 7121
QY 1321 GAATGGGCTGCAAAAGCTGCACCTTCTGGACCTTCTTCATCGGTGCGAAGCTCACGTTCTTC 1380
Db 7122 CAAACCTACGCCAAATCCATTTCACTATCATTAATTCATCGGTAATCTAACTTCTTC 7181
QY 1381 CGCGACACTTCTCGGAGCTGAGGATGCGGCGCTTACATGATATCCCGAAGCC 1440
Db 7182 CCACAACACTTCTCGGCTATCCGGAATGCCCGACCTTACTCGGACTACCCGATGA 7241
QY 1441 TTTGGGCTGTGAAACAAAGTCTCTGCTATGCTGCTTCTGGCTTCCGCTCGTTCCTG 1500
Db 7242 TACACCACATGAACATCCTATCATCTGTAGGCTCATTCATTTCTCTAACAGCAGTAATA 7301
QY 1501 TTCTTCATCGTGATCTTT 1518
Db 7302 TTAATAATTTTCATGATT 7319

RESULT 13

ADP43327

ID ADD43327 standard; DNA; 16569 BP.

XX AC ADD43327;

XX DT 15-JAN-2004 (first entry)

XX Human mitochondrial DNA (mtDNA) SEQ ID NO:501.

DE mitochondrial haplogroup; mitochondrial DNA; mtDNA;

XX single nucleotide polymorphism; SNP; genetic relationship; antidiabetic;

KW neotrophic; neuroprotective; cytosolic; gene therapy; genealogy;

KW forensic; Alzheimer's disease; cancer; type 2 diabetes mellitus; human;

ds.

XX Homo sapiens.

XX WO2003046225-A1.

XX PD 05-JUN-2003.

XX PF 25-NOV-2002; 2002WO-US038276.

XX PR 26-NOV-2001; 2001US-0333622P.

XX PR 28-MAR-2002; 2002US-0369131P.

XX PR 01-APR-2002; 2002US-0369539P.

XX PA (MITO-) MITOKOR.

XX PI Herrnstadt C;

XX DR WPI; 2003-505214/47.

XX Determining single nucleotide polymorphisms in mtDNA or homoplasmic mtDNA

PT mutations, useful for diagnosing and treating diseases, such as

PT Alzheimer's disease, cancer and type 2 diabetes mellitus.

PS Example 2; SEQ ID NO 501; 193pp; English.

XX The present invention describes a method (M1) for determining the

CC mitochondrial haplogroup of a subject, comprising determining in a

CC biological sample with mitochondrial DNA (mtDNA) from a subject, the

CC presence or absence of at least one mitochondrial single nucleotide

CC polymorphism (SNP) that is associated with a mitochondrial haplogroup.

CC Also described: (1) determining a genetic relationship between two

CC subjects; (2) determining a genetic relationship between an unknown

CC source or biological subject from which an unidentified sample is

CC obtained, and a known source or biological subject from an identified

CC sample is obtained; and (3) determining the presence of or the risk of

CC having a disease associated with a mtDNA SNP. Mitochondrial DNA can have

CC antidiabetic, neotrophic, neuroprotective and cytosolic activities, and

CC can be used in gene therapy. M1 and compositions of the present invention

CC are useful for detecting the presence or risk of diseases, treating such

CC diseases, determining the haplogroup of an individual, and establishing
CC genetic relationships between individuals for genealogical and forensic
CC purposes. The diseases include Alzheimer's disease, cancer and type 2
CC diabetes mellitus. The present sequence represents a full length human
CC mtDNA sequence, which is used in the exemplification of the present
XX invention.
SQ Sequence 16569 BP; 5118 A; 5184 C; 2175 G; 4091 T; 0 U; 1 Other;

Query Match 23.8%; Score 398.8; DB 10; Length 16569;
Best Local Similarity 58.2%; Pred. No. 4.6e-96;
Matches 744; Conservative 0; Mismatches 522; Indels 12; Gaps 2;

QY 244 AACGGACACCTGTGGAAGCTCATGTTTACCTACCATGTTATTCGATGATTTCTTTGTG 303
DB 6051 AACGACACATCTACACGTTTATCGTCACAGCCCATGTTTAAATAATCTTTCATA 6110
QY 304 GGTATCCCGGATGTTTGGTGGTGTGTAATCTGATCGCGCTCAATCGGCGCT 363
DB 6111 GTAATACCATCATTAATCGGAGGCTTTGGCAACTGACTAGTTCCTCCCTAATATCGTGCC 6170
QY 364 CCGGATATGGCCTTCCCGGATGATGAACACCTGTCTGTCTGTGCTTTCATTGCCGTACC 423
DB 6171 CCGATATGGCGTTTCCCGGATGAACACATGAAGCTTCTGACTCTTACCTCCCTCTCTC 6230
QY 424 GCGATGGGCGTGGTGGTGTGTCGACCGGGCGGTGACGTCAGCTGGTTCGGCGGT 483
DB 6231 CTACTCTCTCGCATCTCTTAGTGGAGGCGGAGCAGGAACAGGTTGAACATC--- 6287
QY 484 GGTGGTGTCTGTACCGCGCGTGTGACCCCGAGCTGGCTATTTCGATGACCTCGCG 543
DB 6288 -----TACCTCCCTTACAGGAACTACTCCACCTGGAGCTCGGTAGACCTAAC 6341
QY 544 ATTTTCGGGTTTCACTGTGCGGTGCTCTCGATGATGGGCGCATCAATGATCAG 603
DB 6342 ATCTTCTCTTACCTAGCAGGTGTCTCTCTATCTTAGGGGCCATCAATTTCTATCA 6401
QY 604 ACCTTCTTGAACATGGCGCGCCCGCATGACGCTGCACAAAGTGGCTGTTCGTTG 663
DB 6402 ACAATATCAATATAAAACCCCTCGCATACCCAAATACAAAGCCCTCTCTGCTGA 6461
QY 664 TCGATCTTTATCACGGCTTGGTGTGATCTGTGGCGCTCCGGTGTGGCTGGTGAATC 723
DB 6462 TCGTCTCTAATACAGAGTCTCTACTCTCTCTATCTCTCCAGTCTGCTGTCGATC 6521
QY 724 ACATGCTGTGACGACGCTTACTTGGGACGACCTTCTCAATCTGCTGGCGGT 783
DB 6522 ACTATACTACTACAGACGCAACCTCAACACCACTTCTTGACCCCGCGAGGAGA 6581
QY 784 GACCGATCTGTACCAACATCTCTGTGTTCTTTGGGACCCCGGAAGTGTACATCATC 843
DB 6582 GACCCATCTTATACCAACACTTATCTGATTTTTCGGTCACTGAGTTTATATCTT 6641
QY 844 ATTCTGCGCGCTTTGSCATCATGACCATGTGCTGACCTTCTC---GAAAAAGCG 900
DB 6642 ATCTACCGGCTTGGGAATAATCTCCCATATTGTAACTTACTCTCGGAAAAAAGAA 6701
QY 901 GTCTTGGTTTACCTGCGGATGTTGATGCAATGTTGGCAATGTTCTGGGCTTGTG 960
DB 6702 CCATTTGGATACATAGTATGTTCTGAGCTATGATATCAATTTGGCTTCTAGGGTTATC 6761
QY 961 GTCTGGGCGCACCATCTACACGTTGGTATGCTGCTGACCCAGCAATCTTCTTCATG 1020
DB 6762 GTGTGACACACCATATATTACAGTAGAATAGACGTAGACACACAGCATATTTTACC 6821
QY 1021 CTGCGCACCATGTGATGTCGGGTCCGACCGGATTAAGATCTTCTGCTGATCGCACG 1080
DB 6822 TCCGCCACCAATATCATGCTATCCCCACCGGCTCAAGTATTTAGCTGACTCGCCACA 6881
QY 1081 ATGTGGGCGCTCGGTGAGTTCAATTCGCGCATGCTCTGGGCTTGGCTTTATGTTTC 1140
DB 6882 CTCCACGGAACAATATGAATGATCTGCTGCAGTCTCTGAGCGCTTAGGATTCATCTTT 6941

QY 1141 CTGTTACCGTGGTGGTGTGACCGGTATCTGTGCGCCCAAGCGGTCTGGACCGTGCA 1200
DB 6942 CTTTTCACCGTAGGTGGCTGACTGGCAATTGATTAGCAAACTCATCATAGACATCGTA 7001
QY 1201 TATCAGACACCTATTACGTGTGGCGCACTTCCCATTTATGTGATGTGCTGGTGGATC 1260
DB 7002 CTACACGACGCTACTACGTTGTAGCCCACTTCCCACTATGTCTCTATCAATAGGAGTGTA 7061
QY 1261 TTTGGGATCTTCCCGGTATCTACTTTTACATGCCGAAGTTCTCGGCGCGCTTTCCCG 1320
DB 7062 TTTGGCATCATAGAGGCTTCTACTGATTTTCCCTTATTTCTCAGGCTACACCTTAGAC 7121
QY 1321 GAATGGGCTCAAGCTGCACCTTCTGACCTTCTCATCGGTGCGAAGCTCAGTTTCTTC 1380
DB 7122 CAAACCTACGCCAAATCAATTTCTACTATCAATTTCTCGGCGTAAATCTAACTTTCTTC 7181
QY 1381 CCGCAGCATTCTCGGAGCTCAGGATGATGCCCGCGCTTACATGACTATCCGGAAGC 1440
DB 7182 CCAACACATTTCTCGGCTTATCCGAATGCCCGACGTTACTCGGACTACCCCGATGCA 7241
QY 1441 TTGCGGCTGTGGAAACAAAGTCTGCTATGTGTGCTTCTGCGCTTGGCTGCTTCTCCTG 1500
DB 7242 TACCCACATGAACATCTCTATCATCTGTAGGCTCATTTCTCTAACAGCAGTAATA 7301
QY 1501 TTCTTCATCGTATCTTT 1518
DB 7302 TTAATAATTTTCATGATT 7319
RESULT 14
ADD43295
ID ADD43295 standard; DNA; 16569 BP.
XX
AC ADD43295;
XX
DT 15-JAN-2004 (first entry)
XX
DE Human mitochondrial DNA (mtDNA) SEQ ID NO:469.
XX
KW mitochondrial haplogroup; mitochondrial DNA; mtDNA;
KW single nucleotide polymorphism; SNP; genetic relationship; antidiabetic;
KW neotropic; neuroprotective; cytostatic; gene therapy; genealogy;
KW forensic; Alzheimer's disease; cancer; type 2 diabetes mellitus; human;
XX ds.
OS Homo sapiens.
XX
PN WO2003046225-A1.
XX
PD 05-JUN-2003.
XX
PF 25-NOV-2002; 2002WO-US038276.
XX
PR 26-NOV-2001; 2001US-0333622P.
PR 28-MAR-2002; 2002US-0369131P.
PR 01-APR-2002; 2002US-0369539P.
XX
XX (MITO-) MITOKOR.
XX
XX Herrnsstadt C;
XX
XX WPI; 2003-505214/47.
XX
XX Determining single nucleotide polymorphisms in mtDNA or homoplasmic mtDNA
XX mutations, useful for diagnosing and treating diseases, such as
XX Alzheimer's disease, cancer and type 2 diabetes mellitus.
XX
XX Example 2; SEQ ID NO 469; 193pp; English.
XX
XX The present invention describes a method (M1) for determining the
XX mitochondrial haplogroup of a subject, comprising determining in a
XX biological sample with mitochondrial DNA (mtDNA) from a subject, the
XX presence or absence of at least one mitochondrial single nucleotide

CC polymorphism (SNP) that is associated with a mitochondrial haplogroup.
 CC Also described: (1) determining a genetic relationship between two
 CC subjects; (2) determining a genetic relationship between an unknown
 CC source or biological subject from which an unidentified sample is
 CC obtained, and a known source or biological subject from an identified
 CC sample is obtained; and (3) determining the presence of or the risk of
 CC having a disease associated with a mtDNA SNP. Mitochondrial DNA can have
 CC antidiabetic, neurotropic, neuroprotective and cytoskeletal activities, and
 CC can be used in gene therapy. MI and compositions of the present invention
 CC are useful for detecting the presence or risk of diseases, treating such
 CC diseases, determining the haplogroup of an individual, and establishing
 CC genetic relationships between individuals for genealogical and forensic
 CC purposes. The diseases include Alzheimer's disease, cancer and type 2
 CC diabetes mellitus. The present sequence represents a full length human
 CC mtDNA sequence, which is used in the exemplification of the present
 CC invention.
 XX
 SQ Sequence 16569 BP; 5118 A; 5180 C; 2175 G; 4095 T; 0 U; 1 Other;
 Query Match 23.8%; Score 398.8; DB 10; Length 16569;
 Best Local Similarity 58.2%; Pred. No. 4.6e-96;
 Matches 744; Conservative 0; Mismatches 522; Indels 12; Gaps 2;
 244 AACGACACCTGTGGACGTCATGGTTACCTACCATGGTATTCGTGATGATGTTTGTG 303
 6051 AACGACACCATCTACAGGTTATCGTCACAGCCCATGCAATTTGTAATAATCTTCTTCATA 6110
 304 GGTATCCCGCATTTGTCGGTGGTTTGGTAACATCTGATCGCGTGCAGAAATCGCGCT 363
 6111 GTATATCCCATCAATTCGGGCTTTGGCACTGACTAGTTCCCTTAATAATTCGGTGCC 6170
 364 CCGGATATGGCTTCCCGGATGAAACACCTGTCTGTCTGGTGTTCATTTGCGGTATCC 423
 6171 CCCGATATGGGTTTCCCGCATAAACACATAAGCTTCTGACTCTTACCTCCCTCTCTC 6230
 424 CGGATGGCGGTTCGCTGTCGACCGGGGGTACCGTCAGTGGGTTCCGGCGTT 483
 6231 CTACTCTCTGCTGATCTGCTATAGTGGAGCGGAGCAGAAACAGTTGAACAGTCTC 6287
 484 GGTGGTCTGTACCGCGCGCTGTGACCGCGGAAGCTGGCTATTCGATGGACCTCGCG 543
 6288 -----TACCTCCCTTAGCAGGAACTACTCCACCCCTGGAGCCTCCGTAGACCTAAC 6341
 544 ATTTCGCGGTTCACTTGTGGTGGCTCTCGATCATGGGCGCGATCAACATGATCAG 603
 6342 ATCTTCTCTTACACCTAGCAGGTGTCTCTCTATCTTAGGGCCCATCAATTTTCATCACA 6401
 604 ACCTTCTTGAACATGCGCGCGCGCATGACCTGCAAAAGTGGCTTGTCTCTGTCG 663
 6402 ACAATTATCAATAAACCCTGCGCATPAACCAATACCAACGCGCTCTCTGCTCA 6461
 664 TCCATCTTATACGGTTGGCTGATCTCTGCTGGCGCTGCGGTTCTGCTGGTGAATC 723
 6462 TCGTCTCTAATACAGCAGTCTTACTTCTCTATCTCTCCAGTCTCAGTCTGCGATC 6521
 724 ACCATGCTGCTGACCGACCGTAACTTTCGCGACGACCTTCTCAATCTCTGCGCGGT 783
 6522 ACTATACTATACAGACCGACCTTCAACACACCTTCTTGACCCCGGAGGAGA 6581
 784 GACCGGATCTGTACCAACATCTCTGTGTTTCTTGGGACCCCGAAGTGTACATCATC 843
 6582 GACCCCATCTATACCAACACCTATTCTGATTTTTCGGTCACTTGAAGTTTATATCTT 6641
 844 ATTCTGCGGCTTTCGCTATCATCAGCATGCTGCTGCTGACCTTCTC---GAAAGCGG 900
 6642 ATCTACCGGCTTCGGAATAATCTCCCATATGTGAACCTTACTCTCGGAAAAAAGAA 6701
 901 GTCTCGGTACTCGCGATGTTCTATGCAATGGTGGCAATCGGTGTTCTGCGGCTTGTG 960
 6702 CCATTTGGATACATAGTATGGTCTGAGCTATGATATCAATTTGGTCTCTAGGGTTATC 6761
 961 GTCTGGGCGCACATGTACACCGTTGGTATGTGCTGACCCAGCAATCTTACTTATG 1020

Db 6762 GTGTGAGCACCATATATTTTACAGTAGGAATAGAGTAGACACAGAGCATATTTTACC 6821
 QY 1021 CTGGCCACCATGGTATCGCGTGGCGGACCGGCATTAAGATCTTCTCGTGGATCCGACG 1080
 Db 6822 TCCGCTACCATATCATCGTATATCCACCGCGCTCAAGTATTTAGTGTACTGCCACA 6881
 QY 1081 ATGTGGGGCGCTCGGTTGAGTTCAAAATGCGCGATGCTCTGGGCTTTGGCTTTATGTT 1140
 Db 6882 CTCCACGGAAGCAATATGAATGATCTGTCGAGTCTCTGAGCCCTAGGATTCATCTTT 6941
 QY 1141 CTGTTTACCGTGGTGGTGTGACCGGTATCGTGGCGGCGGCTCTGACCGTCA 1200
 Db 6942 CTTTTACCGTGGTGGCTGACTGGCAATGTTATGCAAACTCATCACTAGACATCGTA 7001
 QY 1201 TATCAGCACCATATACGTGGTGGCGCATCTTCCATTAATGATGTCGCTGGTGGATC 1260
 Db 7002 CTACACGACACGTACTAGCTTGTGGCCCATCTCCACTATGCTATCAATAGGAGCTGA 7061
 QY 1261 TTTGGATCTTCCCGGTATCTACTTTTACATGCCGAAGTTCTCGGCGCGCTTTCCCG 1320
 Db 7062 TTGGCATCATAGGAGGCTTCACTCACTGATTTTCCCTTATCTCAGGCTACACCTAGAC 7121
 QY 1321 GAATGGGCTGCAAGCTGCACCTTCTGGACCTTCTTCATCGGTGGGAACGTCACGTTCTTC 1380
 Db 7122 CAAACCTACGCCAAATCCATTTCACTATATATTCATCGGCGTAATCTAACTTTCTTC 7181
 QY 1381 CCGGACGACTTCTGGGACGTCAGGATATCGGCGCGCTTACATCGACTATCCCGAAGCC 1440
 Db 7182 CCACAACACTTCTCGGCTATFCGGAATGCCCGAGCTTACTCGGACTACCCCGATGCA 7241
 QY 1441 TTCGCGCTGTGGAAACAAAGTCTCGTCTATGCTGGTTCCTGCGCTTCCGCTGTTCTG 1500
 Db 7242 TACACCATGAACATCTCTATCATCTGTAGGCTCATTTCTTCTTAACAGCAGTAATA 7301
 QY 1501 TTTCTCATCGTATCTTT 1518
 Db 7302 TTAATAATTTTCATGATT 7319
 RESULT 15
 ADD43302
 ID ADD43302 standard; DNA; 16569 BP.
 XX
 AC ADD43302;
 XX
 DT 15-JAN-2004 (first entry)
 XX
 DE Human mitochondrial DNA (mtDNA) SEQ ID NO:476.
 XX
 KW mitochondrial haplogroup; mitochondrial DNA; mtDNA;
 KW single nucleotide polymorphism; SNP; genetic relationship; antidiabetic;
 KW neurotropic; neuroprotective; cytoskeletal; gene therapy; genealogy;
 KW forensic; Alzheimer's disease; cancer; type 2 diabetes mellitus; human;
 KW ds.
 XX
 OS Homo sapiens.
 XX
 PN WO2003046225-A1.
 XX
 PD 05-JUN-2003.
 XX
 PF 25-NOV-2002; 2002WO-US038276.
 XX
 PR 26-NOV-2001; 2001US-0333622P.
 PR 28-MAR-2002; 2002US-0369131P.
 PR 01-APR-2002; 2002US-0369539P.
 XX
 PA (MITO-) MITOKOR.
 XX
 PI Herrnstadt C;
 XX
 XX WPI; 2003-505214/47.
 XX

Search completed: October 22, 2004, 14:49:47
Job time : 1007 secs

678 GCCTTGGCTGATCTCTGGCGCTGCGGTTCTGGCTGGTGCAATCACCATTGCTGTGAC 737
Db |||
379 CGCGTTCTGCTGATCGCGGTGATGCGGTGCTGGCGCGGTGATGACCATGATGCTGAT 438
Qy |||
738 CGACCGTAACCTCGGACGACCTTCTTCAATCTGCTGGCGCGGTGACCCGATTTCTGTA 797
Db |||
439 GGACATCCACTTCGGACCAAGCTCTTTCAGCGCGCGCGCGCGCGCGCGCTGTT 498
Qy |||
798 CCACACATCTGCTGCTTCTGGGACCGGAGTGTACATCATATTTGCGCGGCTT 857
Db |||
499 CCACACATCTGCTGCTTCTGGGACCGGAGTGTACATCATATTTGCGCGGCTT 558
Qy |||
858 TGGCATCATCAGCCATCTGCTGCTGCGACCTTCTGAAAAGCCGCTTCTTGGTTTACCTGOC 917
Db |||
559 CGGTGCGGTGAGTCCGATCATCCGACCTTCGCGCGGCAAGCGCTGTTGGCTACACCTC 618
Qy |||
918 GATGGTCTATGCAATGCTGGCAATCGGTGTTCTGGGCTTGTGCTGCGCGGACACAT 977
Db |||
619 GATGGTCTATGCAATGCTGGCAATCGGTGTTCTGGGCTTGTGCTGCGCGGACACAT 678
Qy |||
978 GTACACCGCTTGGTATGCTGCTGACCCAGCAATCTACTTCACTGCTGCCACCATGCTGAT 1037
Db |||
679 GTTCTGCTGCTGCGATCCGCTGACCGGCGAGCTGTTCTTCACTAGCCACCATGCTGAT 738
Qy |||
1038 CGCGGTGCGGACCGGCAATTAAGATCTTCTGCTGATGCGGACCATGCTGGGCGGCTCGGT 1097
Db |||
739 CGCGGTGCGGACCGGCAATTAAGATCTTCTGCTGATGCGGACCATGCTGGGCGGCTCGGT 798
Qy |||
1098 TGAGTTCAAAATCGCGATGCTGCGGCTTCTGGGCTTGTGCTGCTGCTGCTGCTGCTGCTG 1157
Db |||
799 GACCTTCGACGCGGATGCTGCTGCGGCTTCTGGGCTTGTGCTGCTGCTGCTGCTGCTGCTG 858
Qy |||
1158 TGTGACCGGTATCGTGTGCGGCAAGGCTTCTGGGCTTGTGCTGCTGCTGCTGCTGCTGCTG 1217
Db |||
859 CTTCGCGGATGATGCTGCGGATGCGGCGGCGGCTTCCAGTACACGACCTTACTT 918
Qy |||
1218 CGTGGTGGCGGCTTCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1277
Db |||
919 CGTGGTGGCGGCTTCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 978
Qy |||
1278 TATCTACTTTTACATGCGGAGTCTTCTGGGCTGCGGCTTCTGGGCTGCGGCTGCGGCTG 1337
Db |||
979 GGCTTACTTCTGCTGCGGAGTGGACCGGCGGCGGCTTCCAGTACACGACCTTACTT 1038
Qy |||
1338 GCATCTTGGACCTTCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1397
Db |||
1039 GCATCTTGGAGCTTCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1098
Qy |||
1398 ACCTGAGGATGCTGCGGCGGCTTACATGCACTATCCGAGGCTTCCGCGCTGCGGCAAA 1457
Db |||
1099 CTTGCGGCGATGCGGCGGAGTGGACCGGCGGCGGCTTCCAGTACACGACCTTACTT 1158
Qy |||
1458 AGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1517
Db |||
1159 GGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1215
Qy |||
1518 TGTCTATACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1577
Db |||
1216 GGTCTATACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1272
Qy |||
1578 CGATACGCTGGAATGAGCGCTGCGGCTTCACTGCGGCTGCGGCGGCGGCTGCGGCGGCGG 1637
Db |||
1273 CGAGGCGCTGAGTGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1332
Qy |||
1638 CAGG 1641
Db |||
1333 CGAG 1336

RESULT 2

US-09-252-991A-676/c

: Sequence 676, Application US/09252991A

: Patent No. 6551795

GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

; PRIORITY FILING DATE: 1999-02-18

Db 1126 ACCAGCTGCTGTTCTCTGTTTCATGTCATCAAGTGCATCCGCGGCGCAAGCGGCGCCCT 1185
Qy 1555 CCGAACCCGTTGGGCGAATTCGCGGATACGCTGGAATGGAGCTGCGCATCACCGCTCCG 1614
Db 1186 GCCAAGCCCTGGGACG---GGCGGAGGCTGGAGTGGAGATCCCTCCGCGGCGCC 1242
Qy 1615 GCCACACGTTGGAACGCTGCCAAG 1641
Db 1243 TACCACACCTTCAGCACCGCGCGGAG 1269

RESULT 4
US-09-377-497-3
; Sequence 3, Application US/09377497
; Patent No. 6670119
; GENERAL INFORMATION:
; APPLICANT: YOSHIKAWA, YOSHIE
; APPLICANT: MURAI, HIROYUKI
; APPLICANT: ASADA, KIYOZO
; APPLICANT: HINO, FUMISUGU
; APPLICANT: KATO, IKUNOSHIN
; TITLE OF INVENTION: CANCER-ASSOCIATED GENES
; FILE REFERENCE: 1422-388P
; CURRENT APPLICATION NUMBER: US/09/377,497
; CURRENT FILING DATE: 1999-08-20
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 3
; LENGTH: 1539
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: any n or xaa = unknown
US-09-377-497-3

Query Match 23.7%; Score 397.2; DB 4; Length 1539;
Best Local Similarity 58.1%; Pred. No. 1.3e-102;
Matches 743; Conservative 0; Mismatches 523; Indels 12; Gaps 2;

Qy 244 AACGGACCTGGAACGTCATGTTACCTAGCAGTGTATCTGATGATGTTCTTTGG 303
Db 148 AAGCACACATCACAACGTTATCGTCACAGCCCATGCAATTTGTAATATCTTCTATA 207
Qy 304 GGTATCCCGCAATGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 363
Db 208 GTAATACCATCATATCGGAGGCTTTGGCACTGACTAGTTCCCTTAATATCGGTGCC 267
Qy 364 CCGGATATGCGCTTCGCGGTATGAACACCTGTGTTCTGGCTGTTCATTCGCGGTACC 423
Db 268 CCGGATATGCGCTTCGCGGTATGAACACCTGTGTTCTGGCTGTTCATTCGCGGTACC 327
Qy 424 GCGATGGGCTGCTTCGCTGTTCGCAACCGGCGGTGAGGTCAGTGGTTCGGGCGTT 483
Db 328 CTACTCTGCTCGCATCTGCTATAGTAGAGCGGAGACGAGTTGAACATC--- 384
Qy 484 GGTGGGTTCTGTACCCGCGGTGTGACCCCGGAGTGGCTATTCGATGACCTCGCG 543
Db 385 -----TACCCTCCCTTAGCAGGAACTACTCCACCCCTGGAGCCTCGTAGACCTAAC 438
Qy 544 ATTTTCGGGTTTCACCTGTCGGTGGCTCTGATCATGCGGCGGATCAACATGATCAG 603
Db 439 ATCTCTCTTACCTACCTAGCAGTGTCTCTCTATCTTAGGGGCACTCAATTTCAACA 498
Qy 604 ACTTCTTGAACATGCGCCCGCGCATGAGCTGCACAAAGTCCCTTGTTCGTTGG 663
Db 499 ACAATATCAATATAAACCCCTGCCATAACCCAAATACCAACGCGCCCTCTTCGTCTGA 558
Qy 664 TCGATCTTATACCGCTTGGCTGTATCTCTGCGGCTGCGGCTGCTGGCTGTGTCATC 723
Db 559 TCGGCTTAAATCAGACAGTCTATCTTCTCCATCTCTCCAGTCTAGCTGCTGGCATC 618
Qy 724 ACCATGCTGCTGACCGACCGTAACCTTCGCGACGACTCTTCAATCTTCGCTGGCGGCT 783

Db 619 ACTATACTAAGACAGACCGCAACCTCAACACCACTTCTTCGACCCCGCGGAGGGA 678
Qy 784 GACCCGATTCGTACCAACACATCTCTGGTGTCTTTGGGCAACCGGAAGTGTATCATCATC 843
Db 679 GACCCGATTCGTACCAACACATCTCTGGTGTCTTTGGGCAACCGGAAGTGTATCATCATC 738
Qy 844 ATTCGCGCGGCTTTGGCATCATCAGCCATCTGCTGTGACCTTCTC---GAAAAAGCG 900
Db 739 ATCTTACAGGCTTCGGAATATCTCCCAATATGTAATCTACTACTCGGAAAAAGAA 798
Qy 901 GTCTTCGTTTACTGCGCGATGGTCTATGCAATGFGGCAATCGGTGTCTGGGCTTTGTC 960
Db 799 CCAITTTGATATACATAGGTATGGTCTGAGCTATGATATCAATTTGGCTTCTAGGTTTATC 858
Qy 961 GTCTGGGCGCACCATGTCACACGTTGGTATGCTGACCCAGCAATCTCTGATGCGCACG 1020
Db 859 GTGTGAGCACACCATATATTTACAGTAGGAATAGACGTAGACACAGAGCATATTTCAAC 918
Qy 1021 CTGCGCAACATGTTGATCGCGGTGCGACCGGCAATTAAGATCTTCTGATGATGCGCACG 1080
Db 919 TCCGCTACCATAAATCATCGCTATATCCCAACCGGCTCAAGTATTTAGCTGACTCGCCACA 978
Qy 1081 ATGTGGGCGGCTCGGTTGAGTTCAAAATCGCCGATGCTCTGGGCTTTGGCTTTATGTC 1140
Db 979 CTCCAGGAGCAATATGAATGATCTGTCAGTGTCTCTGAGCCCTAGGATTCATCTTT 1038
Qy 1141 CTGTTTCCCGTGGTGGTGTGACCGGTATCGTGTGCGCAAGCGGTCTGGACCGTGCA 1200
Db 1039 CTTTTCACCGTAGGTGGCTGACTGCGCATTTGATTAGCAAACTCATCACTAGACATCGTA 1098
Qy 1201 TATCAGACACCTTATAGCTGTGGGCGCACTTCCATTTATGATGTCGTGGTGGCATC 1260
Db 1099 CTACACGACAGTACTAGCTTGTAGTCACTTCCATCTATCTCTCAATAGAGGTGTA 1158
Qy 1261 TTTCCGATCTTCGCGGTATCTACTTTTACATCGCAAGTCTTCGGGCGCGCTTTCCCG 1320
Db 1159 TTTGCCATCATAGAGGCTTCATCTAGTATTTCCCTATTTTCAGGCTACACCTAGAC 1218
Qy 1321 GAATGGCTGCAAGTGTCACTTCTGGAACCTTTTCATCGGTGCGAAGCTGACGTTCTTC 1380
Db 1219 CAAACCTACGCCAAATTCATTTCCGTATCATATTCATCGCGTAAATCTAACTTTCTTC 1278
Qy 1381 CCGCAGCACTTCTTCGGAGCTGAGGTATCGCGCGGTACATCGACTATCCGAGGCC 1440
Db 1279 CCACAACTTCTTCGCGCTATCGGAATGCCCGAGCTTCTCGGACTACCCCGATGA 1338
Qy 1441 TTGCGCTGTGGAACAAAGTCTCTCTATGTCGCTTCTGCGCTTCTGCGCTTCTGCTG 1500
Db 1339 TACACCATATGAATATCTCTATCTAGGCTCATCTATCTTCTTAACAGCAGTAATA 1398
Qy 1501 TTCTTCATCTGATCTTT 1518
Db 1399 TTAATAATTTTCAGATT 1416

RESULT 5

US-08-219-842-1
; Sequence 1, Application US/08219842
; Patent No. 5563323
; GENERAL INFORMATION:
; APPLICANT: Parker, W. D.
; APPLICANT: Hernstadt, Corinna
; TITLE OF INVENTION: Diagnostic and Therapeutic Compositions
; TITLE OF INVENTION: for Alzheimer's Disease
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; City: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/219,842
APPLICATION NUMBER: 30-MAR-1994
FILING DATE: 30-MAR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-AG 9504
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1613 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-219-842-1

Query Match 23.7%; Score 397.2; DB 1; Length 1613;
Best Local Similarity 58.1%; Pred. No. 1.4e-102;
Matches 743; Conservative 0; Mismatches 523; Indels 12; Gaps 2;
QY 244 AACGGACACCTGTGGACGTCATGGTTACCTACCATGGTATTCGTGATGATGTTCTTTGG 303
DB 219 AACGACACATCTACACAGTTTCGTCAGAGCCATGATTTGTAATAATCTTTCATA 278
QY 304 GGATCCCGCATGTTTCGGTGTGTTTGGTAACTATCTGATCGCGCTGCAATCGCGGT 363
DB 279 GTAATACCCATCAATACCGGAGCTTTGGCAACTGACTAGTTCCTCCCTAATAATCGGTGC 338
QY 364 CCGGATATGCTTCCCGCTATGACAACTCTGCTTCTGCTGTTGTTTATTCGCGGTACC 423
DB 339 CCCGATATGGGTTTCCCGCATTAACAACTGCTTCTGACTTCTACCTCCCTCTCTC 398
QY 424 CGGATGGCGGTGCTTCTGCTGTTTGGACCGGGGTGACGGTACGCTGGGTTTCGGCGGT 483
DB 399 CTACTCTCTGCTGCTATCTGCTATGATGGAGCGGAGCAGGAACAGGTTGAACAGTC--- 455
QY 484 GGTGGGTTCTGATACCGCGCTGTCGACCGGGAAGCTGGCTATTCGATGACCTCGG 543
DB 456 -----TACCTCTCTAGAGGAACTACTCCACCTGGAGCCTCCGTGACCTAAC 509
QY 544 ATTTTCGGTTCATCTGTCGGGTGCTCTCTGATCATGGGCGCATCAACATGATCAG 603
DB 510 ATCTTCTCTTACACCTAGCAGGTGTCTCTCTATCTTAGGGGCCATCAATTCATCACA 569
QY 604 ACCTTCTGACATGCGCGCCCGGATGACGCTGCACAAAGTGGCTGTTCTGCTGG 663
DB 570 ACAATATCATATAAACCCTTCCCATTAACCAATACAAACGCCCTCTCTGCTGA 629
QY 664 TCGATCTTTATCAGGCTTGGCTGATCTCTGCGCTGCGGCTGCTGCTGCTGCTGCTGCT 723
DB 630 TCCGCTCTAATACAGCAGCTCTACTTCTCTATCTCTCCAGTCTGCTGCTGCTGCTGCT 689
QY 724 ACCATGCTGTCGCGGACCTTCTGCGGACGACCTTCTGATCTGCTGCTGCTGCTGCTGCT 783
DB 690 ACTATCTACTTACAGACCGCACTTCTGACCCCTCTGACCCCGCGGAGGAGA 749
QY 784 GACCGGATCTGTACCAACATCTCTGTTGTTCTTTGGGACCGGAGTGTACATCATC 843
DB 750 GACCCCATCTTATACCAACACCTATCTGATTTTTCGGTCACTGCTGCTGCTGCTGCTGCT 809
QY 844 ATTGCGCGCTTGGCATCATCAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
DB 810 ATCTACAGGCTTCGGAATATCTCCATATTGTAATCTTACTTACTTACTTACTTACTTACTT 869
QY 901 GTCTTCGCTTACCTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960

DB 870 CCATTTGGATACATAGGTATGGTCTGAGCTATGATCAATTTGGATTCCTAGGTTTATC 929
QY 961 GTCTGGGCGCACCATATGTACACCGTTGGTATGTCGCTGACCCAGCAATCTCTACTTCATG 1020
DB 930 GTGTGAGCACCATATATTTACAGTAGAATAGAGTAGACACACAGGACATATTTTACC 989
QY 1021 CTGGCCACCATGTTGATCGCGGTGCGGACCGGCATTAAGATCTTCTGPGATCGCCACG 1080
DB 990 TCCGCTACCATATATCATCTGCTATCCCAACCGGCTCAAAAGTATTTAGCTGACTGCCACA 1049
QY 1081 ATGTGGGGGCTCGGTTGAGTTCAAATGCGCGATGCTCTGGGCTTTGGCTTTATGTTTC 1140
DB 1050 CTCCAGGAGCAATATGAATGATCTGCTGAGTGTCTGAGCCCTAGGATTCATCTTT 1109
QY 1141 CTGTTACCGTGGTGTGTGACCGGTATCGTGTGGGCCAAAGCGGCTTGGACCGTGA 1200
DB 1110 CTTTTACCGTGGTGTGTGACCGGTATCGTGTGGGCCAAAGCTCATCTAGACATCGTA 1169
QY 1201 TATCAGGACACCTATTACGTTGGGCGCACTTCCATTTATGATGCTGCTGGTGGGATC 1260
DB 1170 CTACAGCAGCTACTACGTTGTAGCCCACTTCCATTTCTTATCTATCAATAGGAGCTGA 1229
QY 1261 TTTGCGATCTTCCCGGTATCTACTTTTACATGCGGAAGTTCTCGGCGGCGCTTTCCCG 1320
DB 1230 TTTGCGATCTAGGAGGCTTCTACTGATTTCCCTTATTTCTAGGCTACACCCCTAGAC 1289
QY 1321 GAATGGGCTCAAGCTGCACCTTCTGACCTTCTTATCGGTGCGAAGCTGACGTTCTTC 1380
DB 1290 CAAACCTACGCCAAATCCATTTCTACTATCATATTCGCGGTAAATCTAACTTCTTC 1349
QY 1381 CCGGACACTTCTCGGAGCTGAGGTATGCGGCGGTGATCATGATTCCTGATTCCTGAGCC 1440
DB 1350 CCAACACACTTCTCGGCTATCCGGAATGCGCCGACGTTACTCGGACTACCCCGATGA 1409
QY 1441 TTCGCGTGTGGAACAAAGTCTGCTCTATGTTGCTTCTGCGCTTCTGCTGCTGCTGCTG 1500
DB 1410 TACACACATGAACATCTCTATCATCTGTAGCTCATCTTCTTCTACACAGCTAATA 1469
QY 1501 TTTCTCATGCTGATCTTT 1518
DB 1470 TTAATAATTTTCATGAT 1487

RESULT 6

US-08-451-096-1
; Sequence 1, Application US/08451096
; Patent No. 5760205
; GENERAL INFORMATION:
; APPLICANT: Parker, W. D.
; APPLICANT: Herrin, Corinna
; TITLE OF INVENTION: Diagnostic and Therapeutic Compositions
; TITLE OF INVENTION: for Alzheimer's Disease
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/451,096
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA: US 08/219,842
; APPLICATION NUMBER:
; FILING DATE: 30-MAR-1994
; ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-AG 9504
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1613 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-451-096-1

Query Match 23.7%; Score 397.2; DB 1; Length 1613;
Best Local Similarity 58.1%; Pred. No. 1.4e-102;
Matches 743; Conservative 0; Mismatches 523; Indels 12; Gaps 2;
QY 244 AACGACACCTGTGGAAACGTCATGTTACCTTACCTACCATGTTATCTTGATGATGTTCTTTGTG 303
DB 219 AACGACACATCTACACGTTATCTGTACAGCCATGCAATTTGTAATATCTTTCTCA 278
QY 304 GGTATCCCGCATTTGTTGGTGGTTTGGTAACCTATCTGATGCGGTGCAAAATCGGCGCT 363
DB 279 GTAATACCCATCAATATCGAGGCTTTGGCAACTGACTAGTTCCTCTTAATATCGGTGC 338
QY 364 CCGGATATGGCTTCCCGGCTATGAACAACTGTCTGCTTCTGGTGTCTATTCGCGGTAC 423
DB 339 CCGGATATGGCTTCCCGGCTATGAACAACTGACTAGTTCCTCTTAATATCGGTGC 398
QY 424 GCGATGGGTGGCTTGGCTTGTGGCAACCGGCGGTGACGCTGAGTGGGTTCGGGCGTT 483
DB 399 CTACTCTGCTGCTATGCTATAGTGGAGCGGAGGAGGACAGGTTGACAGTC--- 455
QY 484 GGTGGGTCTGTACCGCGCTGTGACCGCGGAAGCTGGCTATTCGATGAGCTCGG 543
DB 456 -----TACCTCTCTAGCAGGAACTACTCCACCCCTGGAGCTTCGATGACCTAAC 509
QY 544 ATTTTCGGCTTACCTTGTGGTGGCTCTCTCATATGCGGCGGATCAACATGATCAG 603
DB 510 ATCTCTCTTACACCTAGCAGTGTCTCTATCTTAGGGGCACTAATTTCAATCA 569
QY 604 ACCTCTTGAACATGCGGCGGCGGCGGATGAGCTGCAAAAGTGGCTTGTCTGGT 663
DB 570 ACATTTATCAATAAAGCCCGCTGCAATACCAATACCAAGCCCTCTTCTGCTGA 629
QY 664 TCGATCTTTATACGGCTTGGCTGATCTCTGCGGCTGCGGCTTCTGGCTGCTCAATC 723
DB 630 TCGTCTTAATACACGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 689
QY 724 ACCATGCTGTGACGACGCTTACCTTGGGCAAGCTTCTCAATCTCTGCGGCGGT 783
DB 690 ACTATACTACTACAGACGCAACCTCAACACCACTTCTTCGACCCCGGAGGAGGA 749
QY 784 GACCCGATCTGTACCAACACATCTGTGTTCTTTGGGCAACCGGAGTGTACATATC 843
DB 750 GACCCGATCTGTACCAACACATCTGTGTTCTTTGGGCAACCGGAGTGTATATCTT 809
QY 844 ATTCGCGGCTTTGGGCAATACGACCTATGCTGTCGACCTTCTC---GAAAGGCG 900
DB 810 ATTCGCGGCTTTGGGCAATATCTCCCATATTTGTAATCTACTCTCGGAAAGGAA 869
QY 901 GCTCTGGTGTACCTGCGGATGCTATGCAATGTTGGCAATCGGTGTTCTGGGCTTGT 960
DB 870 CCAATTGATATAGTATGCTGAGCTATGATCAATGGAATCTCTAGGCTTATC 929
QY 961 GTCTGGGCGCACCATGTACACCGTGTGGTATGTGCTGACCCGCAATCTCTCTCATG 1020
DB 930 GTGTGAGCACACCATATATTTACATAGGAATAGACGTAGACACGAGCATATTTTACC 989
QY 1021 CTGGGCGCACCATGTGATGCGGTCGCGGATTAAGATCTCTCTGATGATCGGCAAG 1080
DB 990 TCCGCTACCAATATCATGCTATCCCGGCGGTCAAGATTTAGCTACTCGCCACA 1049

QY 1081 ATGTGGGCGGCTCGGTTGAGTTCAATCGCGGATCTCTGGGCTTGGCTTATGTTT 1140
DB 1050 CTCACCGGAAGCAATATGAATGATCTCTGTCAGTGTCTGAGCCCTAGGATTCATCTTT 1109
QY 1141 CTGTTACCGCTGGTGGTGTGACCGGTATCGTGTGGCCCAAGCGGTCTGGACCGTGA 1200
DB 1110 CTTTTCACCGTAGTGGCTGACTGCGCATTTGATTAGCAAACTCATCACTAGACATCTGA 1169
QY 1201 TATCAGCACCTATTAGTGGTGGCGGACCTTCATTTATGATGTCCTGGTGGGATC 1260
DB 1170 CTACAGCACGCTACTAGCTTGTATGCCCACTTCCACTATGTCCTATCAATAGGAGCTGTA 1229
QY 1261 TTTGGATCTTCGCGGTATCTTACTTTTACATGCGGAAGTTCTCGGCGCGCTTTCCCG 1320
DB 1230 TTTGCCATATAGGAGGCTTCACTTACCTGATTTCCCTATTTCTCAGGCTACACCTAGAC 1289
QY 1321 GAATGGGCTGAAAGCTGACCTTCTGGACCTTCTTCATCGGTGGAAAGCTCACCTTTTC 1380
DB 1290 CAACCTAGCGCAAAATCCATTTCACTATCATATTTCACTCGGCGTAAATCTAACTTTCTC 1349
QY 1381 CCGCAGCACTTCTCGGAGCGTCAAGGTATGCGGCGGTATACATCGACTATCCCGAGCC 1440
DB 1350 CCACAGCACTTCTCGGCTATCGGATGCGGCGGATGCGGCTACTCGGACTACCCGATGCA 1409
QY 1441 TCGGCTGTGGAAACAAAGTCTCTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
DB 1410 TACACCAATGAAACATCTATCATCTGAGGCTCAITTCATTTCTCTAACAGCAGTAATA 1469
QY 1501 TTTCTCATCGTATCTTT 1518
DB 1470 TTAATAATTTCAATTT 1487

RESULT 7

US-08-810-599-1
; Sequence 1, Application US/08810599
; Patent No. 5976798
; GENERAL INFORMATION:
; APPLICANT: PARKER, W. Davis
; APPLICANT: HERENSTADT, Corinna
; APPLICANT: GHOSH, Soumitra S.
; APPLICANT: FAHY, Eoin
; TITLE OF INVENTION: Methods for Detecting Mitochondrial Mutations
; TITLE OF INVENTION: Diagnostic for Alzheimer's Disease and Methods for Determining
; TITLE OF INVENTION: of Mitochondrial Nucleic Acid
; NUMBER OF SEQUENCES: 82
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenyon & Kenyon
; STREET: 1025 Connecticut Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.25" Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 6.1 for Windows
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 08/810,599
; FILING DATE: Concurrent Herewith
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/757,438
; FILING DATE: 27 No. 5976798 1996
; APPLICATION NUMBER: US 08/614,072
; FILING DATE: 12 Mar 1996
; APPLICATION NUMBER: US 08/536,036
; FILING DATE: 29 Sep 1995
; APPLICATION NUMBER: US 08/414,969
; FILING DATE: 31 Mar 1995
; APPLICATION NUMBER: US 08/413,740

FILING DATE: 30 Mar 1995
APPLICATION NUMBER: US 08/410,658
FILING DATE: 24 MARCH 1995
APPLICATION NUMBER: US 08/397,808
FILING DATE: 3 Mar 1995
APPLICATION NUMBER: US 08/219,842
FILING DATE: 30 MARCH 1994
ATTORNEY/AGENT INFORMATION:
NAME: Toifenetti, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 2105/17
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1613 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
HYPOTHETICAL: No
ANTI-SENSE: No
US-08-810-599-1

Query Match 23.7%; Score 397.2; DB 2; Length 1613;
Best Local Similarity 59.1%; Pred. No. 1.4e-102;
Matches 743; Conservative 0; Mismatches 523; Indels 12; Gaps 2;

Qy 244 AAGGACACCTGTGGACGTCATGGTGTACCTACCATGGTATCTGATGATGTTCTTTGGT 303
Db |||||
Qy 219 AAGGACACATCTACAAAGTATCGTCAGACCCATGCAATTTGTAATAATCTTTCATA 278
Db |||||
Qy 304 GGTATCCCGCATGTTTCGGTGTGTTTGGTAACTATCTGATGCGCGTGCMAATCGGCGCT 363
Db |||||
Qy 279 GTAATACCATATAATCGGAGCTTTGGCACTGACTAGTTCCCTTAATATCGTGGC 338
Db |||||
Qy 364 CCGATATGCGCTTCCGGGTATGACACCTCTGCTTCTGGCTGTTTATTCGCGGTACC 423
Db |||||
Qy 339 CCGATATGCGCTTCCCGCATATAACATAAGCTTCTGACTCTTACCTCCCTCTCTC 398
Db |||||
Qy 424 GCGATGGCGGTGGCTTCGCTGTTTCGCACCGGGGGTGCAGGTGAGTTCGGCGGT 483
Db |||||
Qy 399 CTACTCTCTCGCATCTGCTATAGTGGAGCGGAGCAGGATGGAACATC 455
Db |||||
Qy 484 GGTGGGTTCTGTACCGCGCGCTGTGACCCGGAAGTGGCTATTCGATGACCTCGCG 543
Db |||||
Qy 456 -----TACCCTCCCTTAGCAGGAACTACTCCACCCCTGGAGCCTCCGTAGACCTAAC 509
Db |||||
Qy 544 ATTTCGCGGTTCACCTTTCGGGTGCTCTCTGATCATGGCGCATCAACATGATCAG 603
Db |||||
Qy 510 ATCTTCTCTTACACCTAGCAGGTGCTCTCTATCTTAGGGGCCATCAATTCATCACA 569
Db |||||
Qy 604 ACCTTCTTGAACATGCGCGCCCGGCGATGACGCTGCACAAAGTGGCTTCTTCGTGG 663
Db |||||
Qy 570 ACAATTATCAATATAAACCCCTGCGCATAAACCCCAATACCAACGCGCTCTTCGTCTGA 629
Db |||||
Qy 664 TCGATCTTTATCAGGCTTGGCTGTGCTCTGCGCGCTCGGCTTCTGGCTGGTGAATC 723
Db |||||
Qy 630 TCCGTCTCTATCAGCAGTCTTACTTCTCTATCTCTCCAGTCTGCTAGCTGGCATC 689
Db |||||
Qy 724 ACATGCTGTGACCGACGCTTACTTCGGCAAGCTTCTTCAATCTCTGCTGGCGGGT 783
Db |||||
Qy 690 ACTATACTACTAACAGACCGCAACCTCAACACCCCTTCTGACCCCGCGGAGGAGA 749
Db |||||
Qy 784 GACCCGATTCGTACCAACACATCTCTGTGGTCTTTGGGACCGGAAAGTACATCATC 843
Db |||||
Qy 750 GACCCCATCTATACCGACACCTTATCTGATTTTTCGGTACCCCTGAATTTATTTCT 809
Db |||||
Qy 844 ATTCTGCGCGGCTTGGCATCATCAGCCATGTCGTGTCGACCTTCTC---GAAAAGCGG 900
Db |||||
Qy 810 ATCTACAGGCTTCGGAATAATCTCCCATATTTGTAATCTACTCTCCGGAAGAAAGAA 869
Db |||||

Qy 901 GTCTTCGGTTACCTGCGGATGGTCTATGCAATGTTGGCAATCGTGTCTTCTGGGCTTTGTC 960
Db |||||
Qy 870 CCATTTGGATACATAGGTATGCTCGAGCTATGATCAATTTGGATTCCTAGGGTTTATC 929
Db |||||
Qy 961 GTCTGGCGGCACACATGTACACCGTTGCTATGCTGAGCCACCAATCTCTACTTCTCATG 1020
Db |||||
Qy 930 GTGTGAGCACACCATATATTTACAGTAGGAATAGACGTAGACACAGGACATATTTTACC 989
Db |||||
Qy 1021 CTGGCCACCATGATGATCGCGGTGCGGACCGGCATTAAAGATCTTCTCGTGGATCGCCACG 1080
Db |||||
Qy 990 TCCGCTACCATATATCATGCTATCCCAACCGGCGTCAAAGTATTTAGCTGACTCGCCACA 1049
Db |||||
Qy 1081 ATGTGGGGCGGCTCGGTTGAGTTCAAATCGCCGATGCTCTGGGCTTTGGCTTATGTTTC 1140
Db |||||
Qy 1050 CTCACGGAAGCAATATGAATGATCTGCTGAGTGCTCTGAGCCCTAGGATTTTATCTTT 1109
Db |||||
Qy 1141 CTGTTACCGTGGGTGCTGTGACCGGTATCGGCTGCGCCCAAGCGGCTCTGGACCGTACA 1200
Db |||||
Qy 1110 CTTTTCACCGTAGGTGGCGCTGACTGGCAATGTATTAGCAAACTCATCACTAGACATCGTA 1169
Db |||||
Qy 1201 TATCAGGACACCTATTACGTTGGTGGCGGACCTTCCATTTATGATGCTGCTGGTGGATC 1260
Db |||||
Qy 1170 CTACAGGACGCTACTACGTTGTAGCCCACTTCCACTATGCTCTATCAATAGGAGCTGTA 1229
Db |||||
Qy 1261 TTTGCGATCTTCCCGGTATCTACTTTTACATGCCGAAAGTCTCTGGGCGGCGCTTTCCCG 1320
Db |||||
Qy 1230 TTTGCCATCATAGGAGCTTCACTCACTGATTTCCCTATTCTCAGGCTACACCCCTAGAC 1289
Db |||||
Qy 1321 GAATGGCTGCAAGCTGCACCTTCTGGACCTTCTTCATCGGTGCGGACGCTGACGTTCTTC 1380
Db |||||
Qy 1290 CAACTACGCCAAATCCATTTCACTATCATATTCGCGGTAAATCTAACTTTCTTCTC 1349
Db |||||
Qy 1381 CCGCAGCATTCTCTGGAGCTCAGGATGTCGCGCGCGCTTACATCGACTATCCGGAAGCC 1440
Db |||||
Qy 1350 CCAACACATTTCTCGGCTATCCGGAATGCCCGACGTTACTCGGACTACCCCGATGCA 1409
Db |||||
Qy 1441 TTCGCGCTGGGACAAAGCTCTGCTCTATGTTGGTGGTCTGCGCTTCCGCTGCTTCTG 1500
Db |||||
Qy 1410 TACACACATGAACATCTCTATCATCTGTAGGCTCATTTCTTCTAAACAGCAGTAATA 1469
Db |||||
Qy 1501 TTTCTCATCTGATCTTT 1518
Db |||||
Qy 1470 TTAATAATTTATGATT 1487
Db |||||

RESULT 8

US-09-525-906-1
; Sequence 1, Application US/09525906
; Patent No. 6605433
; GENERAL INFORMATION:
; APPLICANT: Jen, Jen
; APPLICANT: Sidransky, David
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Vogelstein, Bert
; APPLICANT: Fliss, Makiko
; APPLICANT: Polyak, Kornelia
; TITLE OF INVENTION: Mitochondrial Dosimeter
; FILE REFERENCE: 1107.85815
; CURRENT APPLICATION NUMBER: US/09/525,906
; CURRENT FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: 09/377,856
; PRIOR FILING DATE: 1999-08-20
; PRIOR APPLICATION NUMBER: 60/097,307
; PRIOR FILING DATE: 1998-08-20
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 16568
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-525-906-1
Query Match 23.7%; Score 397.2; DB 4; Length 16568;

QY	364	CCG	GAT	ATG	GCCTTCCGCGGTATGAA	CAACCTGTCGTTCTGGCTGTTTCATTTGCGGTACC	423
DB	6171	CCC	GAT	ATG	CGCTTTC	CCCGCATAAACAATAAGCTTCTGACTTACCTCCCTCTCTC	6230
QY	424	CGC	ATGGGCGTGGCTTCGCTTTCG	CACCGGCGGTGACGGT	CAGCTCAGCTGGGTTCCGGGGT	483	
DB	6231	CTA	CTCTGCTGCATCTGCTATAG	TGGAGCGCGGACAGAA	CAGGTTGAA	CAGTCTC	6287
QY	484	GGT	TGGGTTCTGTACCCGCGCTGT	CACCGCGGAAGCTGGCTATT	TCGATGGACCTCGCG	543	
DB	6288	----	TAC	CTCCCTT	TAGCAGGAACTACTCCACCCT	CGAGCCTCCGTAGACCTAAACC	6341
QY	544	ATTTT	CGGGTTACATTGTCGGGTGCTCT	CGATCATGGCGGGATCAACATGATCAG	603		
DB	6342	ATCTTCTCTTACACCTAG	CAGGTGCTCTCTATCTTAG	GGGCCATCAATTTCA	TACACA	6401	
QY	604	ACCTTCTTGAA	CATCGCGCCCCCGGATGACG	GTGCACAAAGTGCGGTTGTTCTCGTGG	663		
DB	6402	ACA	ATTATCAATATAAAACCCCTG	CGCATTAACCAATACCAACAGCCCTCTTCGCTCTGA	6461		
QY	664	TGATCTTTAT	CAGGTTGGTGATCTGCTGGCGCTGCGGTTCTGGGCTGGTGAATC	723			
DB	6462	TCCG	TCTAATACAGCAGTCTCTACTTCTCTCTATCTCTCCAGTCTCTAGCTGTGGCATC	6521			
QY	724	ACC	ATGCTGTGACCGACCGTAACTTCGG	CACACACTTCTTCAATCTGCTGGCGCGCGT	783		
DB	6522	ACT	ATACTAA	CAGACCGCACTCAACACACCTTCTTCGACC	CCGCGGAGGGA	6581	
QY	784	GAC	CGGATCTGTACAAACATCTCTGTGTTCTTTGGG	CACCGGAGTGTACATCATC	843		
DB	6582	GA	CCCCATCTATACCAACACCTATTCTGATTTTTCGGT	CACCTGAAGTTTATATTTCTT	6641		
QY	844	ATTCTG	CGCGCTTTGGC	ATCATCAGCCATGTCTGTCTGACCTTCTCTC	--GAAAAAGCCG	900	
DB	6642	ATC	CTACAGGCTTCGGAATACTCCATATTGTA	ACTTACTCTCGGAAAAAAGAA	6701		
QY	901	GTCTT	CGGTACCTGCCGATGGTCTATGCAATGGT	GGCAATCGGTGTTCTGGGCTTTGTC	960		
DB	6702	CCA	TTTGGATACATPAGTATGGTCTGAGCTATGATATCAAT	TGGCTTCTCTAGGCTTTATC	6761		
QY	961	GTCTG	GGGCACCATGTACACGGTTGGTATGTCTGCTG	ACCCAGCAATCCTACTTCATG	1020		
DB	6762	GTGTG	ACACCATATATTACAGTAGGAATAGACGT	AGACACAGCATATTTCAAC	6821		
QY	1021	CTGG	CCACATGTGATCGCGGTCCGACCGGCATTAAGATCTTCTCGT	GGATCGCCACG	1080		
DB	6822	TCCG	CTACCATPAATCATCGCTATCCCA	CGCGGTCAAGTATTTAGCTGACTCGGCACA	6881		
QY	1081	ATGTG	GGGCGCTCGGTTGAGTTCAAATGGCCGATGCTCTGGG	CGTTTGGCTTTATGTTTC	1140		
DB	6882	CTC	ACGGAACATATGAATGATCTGCTG	CACTGCTGAGCCCTAGGATCATCTTT	6941		
QY	1141	CTGTT	CACCGTGGTGTGTACCGGTATCGTGTGGCCCAAGCGG	GTCTGGACCTGTGA	1200		
DB	6942	CTTTT	CACCGTGGTGGCTGACTGGCATTTGATTAG	CAAACTCATCACTAGACATCGTA	7001		
QY	1201	TAT	CACACACCTATTACGTGGTGGCGACATTCATTATG	ATGATGTGCGTGGGTGGCATC	1260		
DB	7002	CTAC	AGCACGTACTAGTTGTAGCCACTTCCACTATGTCTT	ATCAATAGAGCTGTA	7061		
QY	1261	TTT	GGAATTCGCGGPAATCTATTTT	PACATGCGAAGTTCTCGGGCGGCGTTTCCCG	1320		
DB	7062	TTT	GCAATC	ATAGGAGGCTTCATTCAC	TGATTTCCCTATTTCTCAGGCTACACCTAGAC	7121	
QY	1321	GAAT	GGGCTGAAAGCTGCATTTCTG	GACCTTCTTCATCGGTGGGAACGTCA	CGCTTCTTC	1380	
DB	7122	CAA	ACTAGCCAAATCCATTTTCACTATCATATTCAT	CGCGTAATCTPACTTTCTTC	7181		
QY	1381	CCG	CAGCATCTTCTCGGACGTCA	GGGTATGCGCGCGCTTACATCGACTATCCCGAAGCC	1440		
DB	7182	CCAA	CAACTTCTCGGCTATCCGGAATGCCGAGCTTACTCGGATCTACCGGATGCA	7241			

QY	1441	TTGGCGCTGTGGACAAAGTCTCGTCTATGGTGGCTTCTCGCTCGTTCCTG	1500
Db	7242	TACACACATGAACATCCTATCATCTGTAGGCTATTCTTCTTAACAGCAGTAATA	7301
QY	1501	TTCTTCATCGTGATCTTT	1518
Db	7302	TTAATAATTTTCATGATT	7319

```

RESULT 10
US-09-377-856-1
: Sequence 1, Application US/09377856
: Patent No. 6344322
: GENERAL INFORMATION:
: APPLICANT: Polyak, Kornelia
: APPLICANT: Vogelstein, Bert
: APPLICANT: Kinzler, Kenneth
: TITLE OF INVENTION: Subtle Mitochondrial Mutations as Tumor
: TITLE OF INVENTION: Markers
: FILE REFERENCE: 1107.82346
: CURRENT APPLICATION NUMBER: US/09/377,856
: CURRENT FILING DATE: 1999-08-20
: PRIOR APPLICATION NUMBER: 60/097,307
: PRIOR FILING DATE: 1998-08-20
: NUMBER OF SEQ ID NOS: 1
: SOFTWARE: FastSEQ for Windows Version 3.0
: SEQ ID NO 1
: LENGTH: 16569
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-377-856-1

```

6582 GACCCCATCTTATACCAACCTATCTGATTTTTCGGTCACTGAGTTTATATCTT 6641
844 ATTCTGCCCGCTTGGCATCATCAGCATCTGCTGTCGACCTTCTC---GAAAAGCGG 900
6642 ATCTACAGCGCTTCGGAATATCTCCATATTTAACTTACTACTCCGGAATAAGAA 6701
901 GTCTTCGGTTACCTCGCGATGGTCTATSCAATGGTGGCAATCGGTGTTCTGGGCTTTGTC 960
6702 CCAATTGGATACATAGGTATGGTCTGAGCTATGATATCAATTCGCTTCTCTAGGTTTATC 6761
961 GTCTGGGGGACACATGATACAGCTGGTGGTATGCTGCTGACCCAGCAATCTTACTTCAATG 1020
6762 GTGTGAGCACACCATATATTTACAGTAGGAATAGAGTAGACACAGAGCATATTTTCAAC 6821
1021 CTGGCCACCATGGTATCGCGGTGCCACCGGCAATTAAGATCTTCTGCTGGATCGGCACG 1080
6822 TCCGCTACCAATATCATCGCTATCCCAACCGCGTCAAGTATTTAGCTGACTCGGCACA 6881
1081 ATGTGGGGCGCTCGGTGGATTCAMAATCGCCGATGCTCTGGGCTTTGGCTTTATGTTTC 1140
6882 CTCACGGAAGCAATATGAATGATCTGCTGCACTGCTGAGCCCTAGGATTCATCTTT 6941
1141 CTGTTCAACCGTGGTGGTGAACCGGTATCGTGGTGGCCCAACGGGTCTGGACCGTGA 1200
6942 CTTTTCACCGTAGGTGGCTGACTGGGCTATGATAGCAACTCATCACTAGACATCGTA 7001
1201 TATCAGACACCTATACGTGGTGGGCACTTCCATATGATGATGCTGGTGGTGGATC 1260
7002 CTACACGACACTACTACGTGGTGGGCACTTCCACTATGCTGCTATCAATAGGAGCTGA 7061
1261 TTTCCGATCTTCGCGTACTACTTTTATCATGCGGAAGTCTCGGCGCGCTTTCCCG 1320
7062 TTTGCCATATAGAGGCTTCACTGATTTTCCCTATTTTCCAGCTACACCTAGAC 7121
1321 GAATGGCTGCAAGCTTCTTCTGACCTTCTTATCATCGGTGCAACGCTCAGTTTCTTC 1380
7122 CAACCTACGCAAAATCCATTTCACTATCATATTCATCGCGGTAAATCTAATTTCTTC 7181
1381 CCGCAGACATCTCTGGAGCTGAGGTATCGCGCGGTATCATGCACTATCCGAGCC 1440
7182 CCACACATTTCTCGGCTATCCGGAATGCCCGGAGTTTACTCGGACTACCCGATGA 7241
1441 TTCGCGCTGGGAACAAAGCTCGTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
7242 TACACACATGAACATCCTATCATCTGATGCTCATTTCTCTACAGCAGTAATA 7301
1501 TTTCTGATCGATCTTT 1518
7302 TTAATAATTTTCATGATT 7319

RESULT 11

US-09-302-681-2
; Sequence 2, Application US/09302681
; Patent No. 6441149
; GENERAL INFORMATION:
; APPLICANT: Herinstdt, Corrina
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Clevenger, William
; APPLICANT: Faby, Eoin F.
; APPLICANT: Davis, Robert E.
; TITLE OF INVENTION: DIAGNOSTIC METHOD BASED ON
; FILE OF INVENTION: QUANTIFICATION OF EXTRAMITOCHONDRIAL DNA
; FILE REFERENCE: 660088.416C1
; CURRENT APPLICATION NUMBER: US/09/302,681
; CURRENT FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 16569
; TYPE: DNA
; ORGANISM: Homo sapien

US-09-302-681-2
Query Match 23.7%; Score 397.2; DB 4; Length 16569;
Best Local Similarity 58.1%; Pred. No. 4.3e-102; Indels 12; Gaps 2;
Matches 743; Conservative 0; Mismatches 523;
QY 244 AACGACACCTGTGGAGCTCATGTTTACCTACCATGGTATCTGATGATTTCTTTGTG 303
Db AACGACACCTGTGGAGCTCATGTTTACCTACCATGGTATCTGATGATTTCTTTGTG 6110
QY 304 GGTATCCCGGATCTTTCGGTGGTGGTAACTATCTGATGCGCTGCAAAATCGGCGCT 363
Db GGTATCCCGGATCTTTCGGTGGTGGTAACTATCTGATGCGCTGCAAAATCGGCGCT 6170
QY 364 CCGGATATGGCTTCCCGGCTATGAACAACTGCTGCTTCTGCTGCTTCAITGGCGGTACC 423
Db CCGGATATGGCTTCCCGGCTATGAACAACTGCTGCTTCTGCTGCTTCAITGGCGGTACC 6230
QY 424 GGGATGGCGGTGGCTTCCGCTGCTGCGACCGCGGTGACGCTGAGTGGGTTCGGGCGTT 483
Db GGGATGGCGGTGGCTTCCGCTGCTGCGACCGCGGTGACGCTGAGTGGGTTCGGGCGTT 6287
QY 484 GGTGGGTCTGTATCCCGGCTGTCGACCGCGGAGCTGCTATTCGATGGAGCTCGCG 543
Db GGTGGGTCTGTATCCCGGCTGTCGACCGCGGAGCTGCTATTCGATGGAGCTCGCG 6341
QY 544 ATTTTCGGGTCTTCTGCTGGGTGCTCTGCTGATCATGGCGCGCATCAACATGATCAG 603
Db ATTTTCGGGTCTTCTGCTGGGTGCTCTGCTGATCATGGCGCGCATCAACATGATCAG 6401
QY 604 ACCTTCTTGAACATGCGCGCGCGGATGACGCTGCACAAAGTGGTGTCTGCTGG 663
Db ACCTTCTTGAACATGCGCGCGCGGATGACGCTGCACAAAGTGGTGTCTGCTGG 6461
QY 664 TCGATCTTTATCAGCGCTTGGCTGATCTGCTGCGGTGCGGTTCTGGTGGTGGCAATC 723
Db TCGATCTTTATCAGCGCTTGGCTGATCTGCTGCGGTGCGGTTCTGGTGGTGGCAATC 6521
QY 724 ACCATGCTGCTGACCGGCTAACTTCGGCAGCACTTCTTCAATCTGCTGGCGCGGT 783
Db ACCATGCTGCTGACCGGCTAACTTCGGCAGCACTTCTTCAATCTGCTGGCGCGGT 6581
QY 784 GACCCGATCTGTACCAACACATCTGCTGCTTGGGCAACCGGAGGTATCATCATC 843
Db GACCCGATCTGTACCAACACATCTGCTGCTTGGGCAACCGGAGGTATCATCATC 6641
QY 844 ATTTTCGGCGGTGGCATCATCAGCATGCTGCTGCGCTTCTC---GAAAAGCGG 900
Db ATTTTCGGCGGTGGCATCATCAGCATGCTGCTGCGCTTCTC---GAAAAGCGG 6701
QY 901 GTCCTGGGTACCTGCGCGATGGTCTATGCAATGGTGGCAATTCGGTGTCTGGGCTTTGTC 960
Db GTCCTGGGTACCTGCGCGATGGTCTATGCAATGGTGGCAATTCGGTGTCTGGGCTTTGTC 6761
QY 961 GTCTGGGCGCACCATGATACCGGTGGTATGCTGCTGACCCAGCAATCTTACTTCAATG 1020
Db GTCTGGGCGCACCATGATACCGGTGGTATGCTGCTGACCCAGCAATCTTACTTCAATG 6821
QY 1021 CTGGCCACCATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1080
Db CTGGCCACCATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 6881
QY 1081 ATGTGGGCGCGCTCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1140
Db ATGTGGGCGCGCTCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 6941
QY 1141 CTGTTTCACCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1200
Db CTGTTTCACCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 6942
QY 1201 TATCAGACACCTATACGTGGTGGGCACTTCCATATGATGATGCTGGTGGTGGGATC 1260
Db TATCAGACACCTATACGTGGTGGGCACTTCCATATGATGATGCTGGTGGTGGGATC 7001
QY 7002 CTACACGACACTACTAGGTGGGCACTTCCACTATGCTGCTATCAATAGGAGCTGA 7061
Db CTACACGACACTACTAGGTGGGCACTTCCACTATGCTGCTATCAATAGGAGCTGA 7061

QY 1261 TTTCGGATCTTCGGCGGTATCTACTTTTACATGCGGAAGTCTTCGGGCGCGGTTCCTCCG 1320
DB 7062 TTTCGGATCATAGGAGGCTTCATTACCTGAATTCCTCCCTATTCTCAGGCTACACCTAGAC 7121
QY 1321 GAATGGGTGCAAGCTGCATCTTCGGACCTTCTTCATCGGTGCGAAGCTCAGTTCCTTC 1380
DB 7122 CAAACCTACGCGCAAAATCCATTTCACTATCATATTCATCGCGGTAAATCTAACTTTCTTC 7181
QY 1381 CGCAGACCTTCCTCGGACGCTCAGGGTATGCGGCGGTATGCGGCGGTATCATGCACTATCCCGAAGCC 1440
DB 7182 CCACAACTTCTTCGGGCTATCCGGAATGCCGACGTTACTCGGACTACCCGATGCA 7241
QY 1441 TTTCGGGTGGAACAAAGTCTGCTCTATGATGGTGGTTCCTCGGCTTCCTGCTGCTGCTG 1500
DB 7242 TACACCAATGAACATCTATCATCTCTAGGCTCATTCATTTCTCTAACAGCAGTAATA 7301
QY 1501 TTCTTCATCGTATCTTT 1518
DB 7302 TTAATAATTTTCATGATT 7319

RESULT 12
US-09-098-079-2
; Sequence 2, Application US/09098079
; Patent No. 6489095
; GENERAL INFORMATION:
; APPLICANT: Herrnstadt, Corrina
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Cleveland, William
; APPLICANT: Fahy, Eoin F.
; APPLICANT: Davis, Robert E.
; TITLE OF INVENTION: DIAGNOSTIC METHOD BASED ON QUANTIFICATION OF
; TITLE OF INVENTION: EXTRAMITOCHONDRIAL DNA
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSES: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,079
; FILING DATE: 15-JUN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman Ph.D., Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 660088.416
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO. 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16569 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-098-079-2

Query Match 23.7%; Score 397.2; DB 4; Length 16569;
Best Local Similarity 58.1%; Pred. No. 4.3e-102;
Matches 743; Conservative 0; Mismatches 523; Indels 12; Gaps 2;
QY 244 AACGGACCTGTGGAAAGCTAGTGTACCTACCACTGGTATCTGATGATGTTCTTTTG 303
DB 6051 AACGACCACTACACAGTTATCGTCAGCCCAATTCATTGTAATATCTTCATATA 6110

QY 304 GGTATCCCGCATTTGTCGGTGGTTCGTTGTAATACTATCTGATGCGGTGCAAAATCGGCGT 363
DB 6111 GTAATACCCATATAATCGAGGCTTTGGCAACTAGTACTGTTCCCTTAATAATACGTTGCC 6170
QY 364 CCGGATATGCGCTTCGGGATGAAACACCTGTCGTTCTGGCTGTTTCATTTGCGGTACC 423
DB 6171 CCGGATATGCGTTTCGGGATGAAACACATAGCTTCTGACTCTTACCTCCCTCTCTC 6230
QY 424 GCGATGGCGTGGCTTCGCTGTTTCGACCGGGCGGTGACGGTTCAGTGGGTTCGGGCGTT 483
DB 6231 CTACTCTGCTGCACTCTGCTATAGTGGAGCGGAGCAGGAACAGGTTGAACAGTCT-- 6287
QY 484 GGTGGGTTCGTACCGGCGCTGTCGACCGGGAAGCTGGCTATTTCGATGGACCTCGCG 543
DB 6288 -----TACCTCCCTTAGCAGGAACTACTCCACCCCTGGAGCCTCCGTAGACCTAAC 6341
QY 544 ATTTTCGGGTTTCACCTGTCGGGTGCTCTCTGATCATGGGCGCGATCAACATGATCAG 603
DB 6342 ATCTTCTCTTACACCTAGCAGGTGTCCTCTATCTTAGGGGCCATCAATTTTCATCACA 6401
QY 604 ACCTTCTTGAAATGCGGCGCCCGGCGATGACCTGCAACAAAGTGGGTGTTCTGCTGG 663
DB 6402 ACAATTATCAATATAAAACCCCTGCCATAAACCAATACCAACGCGCCCTCTTCGTTCTGA 6461
QY 664 TCGATCTTTATCAGGCTTCGGTGTATCTGCTGGCGTGCCTGCTTCTGGCTGGTGAATC 723
DB 6462 TCGTCTCTATCAGCAGTCTTACTTCTCTATCTCTCCAGTCTCTAGTCTGGCATC 6521
QY 724 ACCATGCTGCTACCGACCGTAACTTTGGCAGCACTTCTTCAATCTGCTGGCGGCGT 783
DB 6522 ACTATACTACTAACAGACCGCAACCTCAACACACCTTCTTCGACCCGCGGAGGAGGA 6581
QY 784 GACCCGATTTCTGTACCAACACATCTCTGTTGTTCTTTGGCACCAGGAGTATCATCATC 843
DB 6582 GACCCGATTTCTATACCAACACCTTCTGATTTTTCGGTCACTCTGAAGTTTATTTCTT 6641
QY 844 ATTTCTGCGCGCTTTTGGCATCATCAGCCATCTGCTGTCGACCTTCTC--GAAAGACCG 900
DB 6642 ATCTACACGCTTCGGAATAATCTCCCAATTTGTAATTTACTACTCCGGAAGAGAA 6701
QY 901 GTCTTGGTTTACCTGCGCATGGTCTATGTAATGTTGGCAATCGGTGTTCTGGGCTTTGTC 960
DB 6702 CCATTTGGATACATAGTATGGTCTGAGCTATGATATCAATTTGGCTTCTTAGGTTTATC 6761
QY 961 GTCTGGGCGACACATGTACACCGTTGTTGCTGCTGACCCAGCAATCTTACTTCTCATG 1020
DB 6762 GTGTGAGCACACCATATATTTACAGTAGGAATAGACGTAGACACACAGCATATTTACCC 6821
QY 1021 CTGCGCAACATGTTGATGCGGCTGCCGACCGGCAATTAAGATCTTCTGCTGATGCGCACG 1080
DB 6822 TCCGCTACCATATCATCGCTATCCCGACCGGCTCAAGTATTTAGCTGACTCGCCACA 6881
QY 1081 ATGTGGGCGCTCGGTTGAGTTCAAAATCGCGATGCTCGGGCTTTGGCTTTATGTTTC 1140
DB 6882 CTCACGGAAGCAATATGAATGATCTGCTGAGTGTCTGAGCCCTAGGATTCATCTTTT 6941
QY 1141 CTGTTACCGGTGGTGGTGTGACCGGTATCGTGTGCGCCCAAGCGGTCTGGACCGTGCA 1200
DB 6942 CTTTTACCGTAGTGGCTGACTGGCAATTTGATTTAGCAAACTCATCACTAGACATGTA 7001
QY 1201 TATCAGACACCTATTAGGTGGCGCACTCCATTTATGATGTCGCTGGGTGGCATC 1260
DB 7002 CTACACGACACGTACTAGTTGTAGCCCACTTCCACTATGTCCTATCAATAGGAGCTGTA 7061
QY 1261 TTTGCGATCTTCGCGCGGTATCTACTTTTATCATGCGGAAGTTCTCGGCGCGGCTTTCCCG 1320
DB 7062 TTTGCCATCATAGGAGGCTTCATTTCACTGATTTCCCTTATTTCTCAGGCTACACCTAGAC 7121
QY 1321 GAATGGGTGCAAGCTGCATCTTCGGACCTTCTTCATCGGTGCGAAGCTCAGCTTCTTC 1380
DB 7122 CAAACCTACGCGCAAAATCCATTTTCACTATCATATTTATCGGCGGTAAATCTAACTTTCTTC 7181


```
; NUMBER OF SEQ ID NOS: 108
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 6691
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(6691)
; OTHER INFORMATION: n = A,T,C or G
US-09-302-681-3

Query Match      23.6%; Score 395.2; DB 4; Length 6691;
Best Local Similarity 57.7%; Pred. No. 1e-101;
Matches 738; Conservative 3; Mismatches 525; Indels 12; Gaps 2;

QY 244 AACGACACCTGTGAGAGTCTGTGTTACCTACCATGTTCTCTGATGATGTTCTTTTGG 303
DB 2582 AACGACACCATCTACAAAGTTATGTCACAGCCCATGATTTGTAATATCTTCTTCATA 2641
QY 304 GGTATCCCCCGCATTTGTCGGTGGTTTGGTAACTATCTGATCCGCTCAAAATCGCGCT 363
DB 2642 GTAATACCCATCAATAATCGAGGYTTTGGCAASTGACTAGTTCCCTAATAATCGGTGCC 2701
QY 364 CGGATATGGCTTCCCGGATGATGAACACCTGTGTTCTGCTGTTCTTCAATGCGGTACC 423
DB 2702 CCGGATATGGGTTTCCCGGATGATGAACACCTGTGTTCTGCTGTTCTTCAATGCGGTACC 2761
QY 424 CGGATGCGGCTGGCTTGGCTGTTTGGCAGCGGCGGTGACCGTCAAGTGGTTCGGGCGTT 483
DB 2762 NTACCTCTGTTNGCATCTCTATAGTGGAGCGGGG-----CAGGACAGGTTGA 2812
QY 484 GGTGGGTTCTGTACCCCGCGGTGTGCGACCCCGGAGCTGGCTATTCGATGACCTCGG 543
DB 2813 ACAGTNTACCCCTCTTGGCAGGAACTACTCCACCCTGGAGCCTCCGTAGACCTAAC 2872
QY 544 ATTTTCGGGTTTCACTTGTGCGGTGCTCTCGATCATGGCGGATCAACATGATCAG 603
DB 2873 ATCTTCTCTTACCTACAGAGTATCTCTCTATCTTAGGACCATCAATTTTCATCA 2932
QY 604 ACCTTCTGAACATGCGCGCCCGCATGACGCTGACAAAGTGGGTTTCTGCTGG 663
DB 2933 ACAATTTAATAAAGCCCTGCGTAAACCCAAATACCAAGCCCTTTTTCGTCGA 2992
QY 664 TCGATCTTTATCAGGCTTGGTGTATCTGCTGGGCTGGGCTGCTGGCTGCTGCTGCTCAATC 723
DB 2993 TCGTCTCTAATCAGAGCTTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3052
QY 724 ACCATGCTGCTGACCGACGTAACCTTGGCAGCACCTTCTTCAATCTCTGCTGCGCGGT 783
DB 3053 ACTATACTACTACAGACCGTAACCTCAACACCACTCTCTCGACCCCGGAGGAGGA 3112
QY 784 GACCGATCTGTACCAACACATCTCTGTTGTTTGGGACCCCGGAGTGTACATCATC 843
DB 3113 GACCCCATTTCTATACCAACACCTATTCTGATTTTTCGGTCACCTGAAATTTATATCTC 3172
QY 844 ATTCTCGCGGCTTGGCATCATCAGCATGTCGTTGTCAGCTTCTC---GAAAGCGG 900
DB 3173 ATCTTACAGGCTTGGATTAATCTCCATATTGTAACCTACTCTCCGAAATAAGNA 3232
QY 901 GTCTTCGGTTTACCTGCGGATGTTGTTATGCAATGGTGGCAATCGGTTTCTGGGCTTTGTC 960
DB 3233 CCATTTGGATACATAGTATGTTCTGAGCTATGATATCAATTTGGCTTCTAGGTTTATC 3292
QY 961 GTCTGGGCGACCATGATACCGGTTGGTATGTCGCTGACCCAGCAATCTTCTCATG 1020
DB 3293 GTGTGAGCACCATATATTTCAGTAGGATAGAGTAGACACAGACATATTTCAC 3352
QY 1021 CTGCCCCACCATGTTGTCGGGTGCGGACCGGATTAAGATCTTCTCGTGGATCGCCAG 1080
DB 3353 TCCGCTACCATATATCTGCTATCTCCCGGCTCAAAAGTATTTAGCTGACTCGCCACA 3412
QY 1081 ATGTGGGCGGCTCGGTTGAGTTCAATCGCCGATGCTCTGGGCTTTGGCTTTATGTT 1140
```

RESULT 15

```
US-09-097-889-1
; Sequence 1, Application US/09097889
; Patent No. 6218117
; GENERAL INFORMATION:
; APPLICANT: Herxstadt, Corrina
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Davis, Robert E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING
; TITLE OF INVENTION: AGENTS THAT QUANTITATIVELY ALTER DETECTABLE
; TITLE OF INVENTION: EXTRAMITOCHONDRIAL DNA: MITOCHONDRIAL DNA RATIOS
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,889
; FILING DATE: 15-JUN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Roseman P.D., Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 660088.417
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6744 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

US-09-097-889-1

Query Match 23.2%; Score 388.2; DB 3; Length 6744;
Best Local Similarity 57.1%; Pred. No. 9.9e-100;
Matches 730; Conservative 19; Mismatches 517; Indels 13; Gaps 3;

```
QY 244 AACGGACACCTGTGGAACGTCATCGTTACCTACCATGTTATCTTGATGATGTTCTTTTG 303
DB |||||
QY 2625 AACGACACATCTACACAGTATCGTCACAGCCATGCATTTGTAATATCTTTTCATA 2684
DB |||||
QY 304 GGTATCCCCGGAATGTTTCGGTGGTTTTGGTAACATCTGATGCCGTGCAAAATCGCGCT 363
DB |||||
QY 2685 GTAATACCCATCAATATCGGAGGCTTTGGCAACTGACTAGTTCCTCCCTAATAATCGGTGCC 2744
DB |||||
QY 364 CCGGATATGGCTTCCCGCGGTATGAACAACTGTCTGGTTCCTGGTTCATTTGCCGCTAAC 423
DB |||||
QY 2745 CCGATATGGCTTCCCGCATMAACAATAGCTTTGACTTTTACCTCCCTCTCTC 2804
DB |||||
QY 424 GCGATGGCGGTGCTTCGCTGTTTCGACCGGGGTGACGTCAGCTGGGTTCGGCGCTT 483
DB |||||
QY 2805 MTACTCCTGTYGCATCTGCTATAGTGGAGCGCGMGCAGAAACAGGTTGAAACAGTM--- 2861
DB |||||
QY 484 GGTGGGTTCTGTACCGCGCGCTGTGACCGCGAGCTGGCTATTCGATGGACCTCGG 543
DB |||||
QY 2862 -----TACCTCCCTTRGAGGAACTACTCCACCMGTGGAGCTCCGCTAGACATAACC 2915
DB |||||
QY 544 ATTTTCGGGTTCACTTGTGCGGTGCTCCG-TGGATCATGGGCGGATCAACATGATCAC 602
DB |||||
QY 2916 ATCTTSCCTTACACYTAGCAGGTCTCTCTCTAATCTTAGGGGCCATCAATTTCAATCAC 2975
DB |||||
QY 603 GACCTTCTTGAACTGCGCGCCCGCGCATGACGTGTCACAAAGTGCCTGTTTCTCGTG 662
DB |||||
QY 2976 AACAAATATYAATAAATAAACCCTCGCATAAACCCCAATACCAACGCCCTTCTGCTG 3035
DB |||||
QY 663 GTCGATCTTTATCACGGCTTGGGTGATCTGCTGGCGTCCCGGTTCTGGCTGGTCAAT 722
DB |||||
QY 3036 ATCCGTCTCTAATCACACAGCTCTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3095
DB |||||
QY 723 CACATGCTGCTGACCGACCGTAATCTGCGGACGACCTTCTTCAATCTGCTGGCGGG 782
DB |||||
QY 3096 CACTATACTACTAACAGACCGYAMCTCAACACACACCTTCTTCTGACCCCTGAAAGTTTATCT 3155
DB |||||
QY 783 TGACCCGATCTGTACCAACACATCTCTGTTCTTTGGCACCCGGAAGTGTACATCAT 842
DB |||||
QY 3156 AGACCCCATCTATACCAACACCTATCTGATTTTTCGGTCACTCTGAAAGTTTATCTCT 3215
DB |||||
QY 843 CATCTCCCGGCTTTGGCATCATGACCCATGTCGTGCGACCTTCTCG---AAAAAGCC 899
DB |||||
QY 3216 YATCCTACACAGGCTTCCGAATAATCTCCCATATGTAACCTTACTCTCCGGAAAAAAGA 3275
DB |||||
QY 900 GGCTTCCTCGGTACTCCGCTGCTCTATGCAATGGTGGCAATCGGTGTTCTGGGCTTTGT 959
DB |||||
QY 3276 ACCATTTGGATACATAGGTATGGTCTGAGCTATGATATCAATTTGGCTTCTAGGGTTTAT 3335
DB |||||
QY 960 CGTCTGGGCGCACCATATGACACCTTTGATGCTGCTGACCCAGCAATCTCTTCTCAT 1019
DB |||||
QY 3336 CGTGTGAGCACACCATATATTTACAGTAGGAATAGACGTAGACACAGAGCATATTTTAC 3395
DB |||||
QY 1020 GCTGGCCACCATGGTATCGCGGTGCGGACCGCATTAAGATCTTCTCGTGGATCGCCAC 1079
DB |||||
QY 3396 CTCGGCTACCATATATCATGCTATCTCCCAACCGCGCTCAAAGTATTTAGCTGACTCGCCAC 3455
DB |||||
QY 1080 GATGTGGGGCGGCTCGGTTGAGTTCAAAATCGCCGATGCTCTGGGCGCTTTGGGCTTTATGTT 1139
DB |||||
QY 3456 ACTCCAGGAAGCAATATGAATGATCTGCTGAGTGTCTCTGAGCCCTAGGATTTATTTT 3515
DB |||||
QY 1140 CTTGTTTACCGTGGGTGTGACCGGTATCGGTGCTGGCCCAAGCGGGTCTGACCGGTGC 1199
DB |||||
QY 3516 TCTTTTACCGTAGGTGGCTGACTGGCATTTGATTAGCAAACTCATCACTAGACATCGT 3575
DB |||||
QY 1200 ATATCAGCACACTATTACGTGTGGCGCACTTCCATTAATGATGCTGCTGGTGGCAT 1259
DB |||||
QY 3576 ACTACAGCACAGTACTACGTTGTAGCCCACTTCCACTATGCTCTATCAATAGGAGCTGT 3635
DB |||||
```

```
QY 1260 CTTTCGATCTTTCGCCGGTATCTACTTTTATCATGCCGAAGTTCTCGGCGCCGCTTTCCC 1319
DB |||||
QY 3636 ATTTGCCATCATAGGAGGCTTCACTCACTGATTTTCCCTATTTCTCAGGGTACACCCCTAGA 3695
DB |||||
QY 1320 GGAATGGGCTGCAAAAGCTGCACCTTCTGGACCTTCTCATCGGTGCGAAAGTCAAGTTCTT 1379
DB |||||
QY 3696 CCAAACTACGCAAAATCGATTTTCGTATCATATTCATCGCGCTAAATCTAACTTTCTT 3755
DB |||||
QY 1380 CCGGACGACCTTCTGGGACGTCAGGGTATGCCGCGGCTTACATCGACTATCCGGAAGC 1439
DB |||||
QY 3756 CCCACAAACACTTCTTCGGCTATCCGGAATGCCCGACGTTACTCGGACTAYCCCGATGC 3815
DB |||||
QY 1440 CTTTCGCGCTGGAACAAAGTCTGCTCTATGCTGCTGCTTCTGCGCTTTCGCTCGTTCTT 1499
DB |||||
QY 3816 ATACACCATGAAAYATCCTATCATCTGAGGTCTCATTTCTCTAACAGCAGTAAT 3875
DB |||||
QY 1500 GTTCTTCACTGATCTTTT 1518
DB |||||
QY 3876 ATTAATAATTTTCATAATT 3894
DB |||||
```

Search completed: October 22, 2004, 12:24:08
Job time : 181 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 22, 2004, 11:44:00 ; Search time 856 Seconds
(without alignments)
10020.378 Million cell updates/sec

Title: US-09-712-768A-1

Perfect score: 1674

Sequence: 1 atggcagacgcgcattca.....acaagcatcctcgcaactaa 1674

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3407233 seqs, 2561960514 residues

Total number of hits satisfying chosen parameters: 6814466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

```
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
5: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	515.6	30.8	1593	15	US-10-127-032-32
2	515.6	30.8	1593	16	Sequence 32, Appl
3	406.2	24.3	1629	16	Sequence 9, Appl
4	398.8	23.8	16559	16	Sequence 25546, A
5	398.8	23.8	16559	16	Sequence 150, Appl
6	398.8	23.8	16559	16	Sequence 145, Appl
7	398.8	23.8	16559	16	Sequence 144, Appl
8	398.8	23.8	16559	16	Sequence 181, Appl
9	398.8	23.8	16559	16	Sequence 140, Appl
10	398.8	23.8	16559	16	Sequence 147, Appl
11	398.8	23.8	16559	16	Sequence 474, Appl
12	398.8	23.8	16559	16	Sequence 476, Appl
					Sequence 501, Appl

```
13 398.8 23.8 16569 16 US-10-308-264-508
14 398.8 23.8 16569 16 US-10-308-264-524
15 398.8 23.8 16569 16 US-10-308-264-529
16 398.8 23.8 16570 16 US-10-308-264-18
17 398.8 23.8 16570 16 US-10-308-264-31
18 398.8 23.8 16570 16 US-10-308-264-101
19 398.8 23.8 16570 16 US-10-308-264-133
20 398.8 23.8 16570 16 US-10-308-264-144
21 398.8 23.8 16570 16 US-10-308-264-147
22 398.8 23.8 16570 16 US-10-308-264-153
23 398.8 23.8 16570 16 US-10-308-264-240
24 398.8 23.8 16570 16 US-10-308-264-285
25 398.8 23.8 16570 16 US-10-308-264-286
26 398.8 23.8 16570 16 US-10-308-264-454
27 398.8 23.8 16571 16 US-10-308-264-26
28 398.8 23.8 16571 16 US-10-308-264-62
29 398.8 23.8 16571 16 US-10-308-264-94
30 398.8 23.8 16571 16 US-10-308-264-205
31 398.8 23.8 16571 16 US-10-308-264-289
32 398.8 23.8 16571 16 US-10-308-264-341
33 398.8 23.8 16572 16 US-10-308-264-54
34 398.8 23.8 16572 16 US-10-308-264-287
35 398.8 23.8 16572 16 US-10-308-264-535
36 398.8 23.8 16573 16 US-10-308-264-452
37 398.8 23.8 16573 16 US-10-308-264-463
38 398.8 23.8 16574 16 US-10-308-264-59
39 398.8 23.8 16574 16 US-10-308-264-430
40 398.4 23.8 2174 13 US-10-027-632-266288
41 398.4 23.8 2174 13 US-10-027-632-266289
42 398.4 23.8 2174 15 US-10-027-632-266288
43 398.4 23.8 2174 15 US-10-027-632-266289
44 398.4 23.8 16568 16 US-10-308-264-382
45 398.4 23.8 16572 16 US-10-308-264-269
```

ALIGNMENTS

```
RESULT 1
US-10-127-032-32
; Sequence 32, Application US/10127032
; Publication No. US20030113742A1
; GENERAL INFORMATION:
; APPLICANT: Whiteley, Marvin
; APPLICANT: Bangera, M. Gita
; APPLICANT: Lory, Stephen
; APPLICANT: Greenberg, Everett Peter
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE MODULATION OF
; FILE OF INVENTION: BIOFILM FORMATION
; FILE REFERENCE: UIZ-070CP
; CURRENT APPLICATION NUMBER: US/10127,032
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,190
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/344,142
; PRIOR FILING DATE: 2001-10-24
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 1593
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-10-127-032-32
```

Query Match 30.8%; Score 515.6; DB 15; Length 1593;
Best Local Similarity 63.0%; Pred. No. 7e-146;
Matches 854; Conservative 0; Mismatches 484; Indels 18; Gaps 3;

```
OY 286 CTGATGATGTTCTTTGGTATCCCGGATGTTTGGTGTGTTGTAATCTATCTGATG 345
Db 247 CTGATCATGTTCTTCGGCGGGTATGCGGCGCTTCGCGCCCTGGGCAACTGGATGATC 306
OY 346 CCGCTGCAAAATCGCGGCTTCGGATATGCGCTTCCCGGCGTATGAAACAACCTGCTGTTCTGG 405
```

Db 307 CCGCTGATGATCGGCGCGGACATGGCCCTCGCGGGATGACAACTTCACTTCTGG 366
Qy 406 CTGTTTCAATGCGGATACCGGATGCGGCTGGCTTGGCTGATGCGGACCGGCGGTACGGT 465
Db 367 CTGTCGCGCGGCGCTTTCGGCTCTGCTGCTGAGCACTTGTTCATGCGCGCGG----- 419
Qy 466 CAGCTGGTTCGGCGCTTGGTGGTCTGTACCGCGCTGTCGACCGCGGAAGCTGGC 525
Db 420 -----CGGCCCCAACTTCGCGTGGACCTTCTATGCGCGCGCTGTCGACCACTTCGCGCCG 474
Qy 526 TATTCGATGACCTTCGGAATTTTCGCGGTTCATGTCGCGGTCTCTCGATCATGCGG 585
Db 475 CACAGCTGACCTTCTTTCATCTTCGCTGACCTTCACCTGCGCGGGATCAGCTCATGCGG 534
Qy 586 GCGATCAACATGACGACCTTCTTGAACATGCGGCGCGCGCGGATGCGTGCACAA 645
Db 535 GCGATCAACGATGACGACCACTTCAACCTGCGCGCGCGCGGATGACCTTCATGAAG 594
Qy 646 GTGCGCTTGTTCGCTGCTGATCTTTATCACGGCTTGGCTGATCTCTGCTGGCGCTGGC 705
Db 595 ATGCGCTGTTCGCTGACCTGCTGATCAACCGCTTCTTTTGTGCGGTGATGCT 654
Qy 706 GTTCTGCTGATGACATGACGCTGCTGACCGACCGTAACTTCGCGACGCTTCTTC 765
Db 655 GTGCTGCGCGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 714
Qy 766 AATCTCTGCGCGGCTGACCGGATCTGTACCAACATCTCTGCTGCTTCTTGGGAC 825
Db 715 AGCGCGCGCGCGCGCGCGCTGCTGCTGACGACGCTGCTGCTGCTGCTGCTGCTGCT 774
Qy 826 CCGGAATGATCATCATCTGCGCGCTTGGCATCATGACCATGCTGCTGCTGCTGCTGCT 885
Db 775 CCGAGGTGATCATCATCTGCGCGCTTGGCATCATGACCATGCTGCTGCTGCTGCTGCT 834
Qy 886 TTCTCGAAAGCGGCTTTCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 945
Db 835 TTGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 894
Qy 946 GTTCTGCGCTTTCGCTGCTGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCT 1005
Db 895 TTCTCTCTCTGCTGCTGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCT 954
Qy 1006 CAATCTCTCTGCTGCTGCGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCT 1065
Db 955 GAGCTGCTTCTGATGACCGCGCGCGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCT 1014
Qy 1066 TCGTGGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCT 1125
Db 1015 AACTGGGTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCTGCTGCT 1074
Qy 1126 TTGGCTTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1185
Db 1075 GTGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1134
Qy 1186 GGTCTGACCGCTGATATGACGACACCTATTAGTGTGGCGCGCGCGCGCGCGCGCGCG 1245
Db 1135 CCGCGAGCTTCAGTACGACGACACCTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1194
Qy 1246 TCGTGGGTGCGATCTTTGCGGATCTTTCGCGGATCTTCTTATGCTGCTGCTGCTGCTGCT 1305
Db 1195 GTGCG 1254
Qy 1306 GCGCGCGCTTTCGCGGATGGGTGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1365
Db 1255 GCGCGATGATGACGAGGACCTTCGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1314
Qy 1366 AAGTCACTGCTTTCGCGGAGCTTCTTGGGAGCTGAGGATGCTGCTGCTGCTGCTGCTGCT 1425
Db 1315 AAGTGGGCTTCTTCCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1374
Qy 1426 GACTATCCGAGGCTTTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1485

Db 1375 GACTACAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1431
Qy 1486 TTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1545
Db 1432 TTCGCGACCAACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1491
Qy 1546 GAGACCGCTGCGAAACCGCTGCGGGAATTCGCGATGCTGCTGCTGCTGCTGCTGCTGCT 1605
Db 1492 CCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1548
Qy 1606 CCGCTGCTGCGGCG 1641
Db 1549 CCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1584

RESULT 2
US-10-389-647-9
; Sequence 9, Application US/10389647
; Publication No. US20040033549A1
; GENERAL INFORMATION:
; APPLICANT: GREENBERG, E. Peter
; APPLICANT: SCHUSTER, Martin
; APPLICANT: LOSTROH, Candi
; TITLE OF INVENTION: QUORUM SENSING SIGNALING IN BACTERIA
; FILE REFERENCE: UIZ-038CP
; CURRENT APPLICATION NUMBER: US/10/389,647
; CURRENT FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: 09/653730
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 60/153022
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 710
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1593
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-10-389-647-9

Query Match 30.8%; Score 515.6; DB 16; Length 1593;
Best Local Similarity 63.0%; Pred. No. 7e-146;
Matches 854; Conservative 0; Mismatches 484; Indels 18; Gaps 3;

Qy 286 CTGATGATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 345
Db 247 CTGATCATGCTTTCGCGCGGCTGATCCGCGCTTCTGCGCGCTGCGCGCGCTGCGCGCG 306
Qy 346 CCGTGCAGAAATCGCGCTGCGGATATGCGCTTCCGCGCTATGAACACCTGTCTGCTGCTG 405
Db 307 CCGCTGATGATCG 366
Qy 406 CTGTTTCAATGCGGATACCGGATGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 465
Db 367 CTGTCGCGCGCGCTTTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 419
Qy 466 CAGCTGCTGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 525
Db 420 -----CGGCCCCAACTTCGCGTGGACCTTCTATGCGCGCGCTGCTGCGACCTTCGCGCCG 474
Qy 526 TATTCGATGACCTTCGCGATTTTCGCGGTTCATGCTGCGGTGCTGCTGCTGCTGCTGCTGCT 585
Db 475 CACAGCTGACCTTCTTTCATCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 534
Qy 586 GCGATCAACATGATGACGACCTTCTGAACATGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 645
Db 535 GCGATCAACGATGATGCG 594
Qy 646 GTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 705
Db 595 ATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 654
Qy 706 GTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 765

Db	655	GTGCTGGCGGCGTGTGTGACCAATGATGCTGATGGACATCACTTCGGACACGACTTCTTC	714
Qy	766	AATCTCTGCTGGCGGCGGTGACCCGATTTCTGTACCAACACATCCTGTGGTCTTTGGGCAC	825
Db	715	AGCGCGCGCGGCGGCGACCCGGTCTGTTCCAGCAGTGTCTGTGTTCTTCGCGCAC	774
Qy	826	CCGGAAGTGTACATCATCTCTGCCGGCTTTTGGCATCATCAGCCATGTCTGTCGAC	885
Db	775	CCGAGGTGTACATCATGATCTTGCCTGCGGCTTCGGTGGGTGAGTGGCATCATCCCGACC	834
Qy	886	TTCTCGAAAAAGCGGCTTCTCGGTTCACCTCCGATGCTTATCGAATGGTGGCAATCGGT	945
Db	835	TTTCGCGCGAAGCGGCTGTTTCGGCTACACCTCGATGCTTACGCCACCGCCAGCATCGCC	894
Qy	946	GTTCTGGGCTTTGTGCTGGGCGACCAATGTACACGTTGGTATGTCTGTGACCCAG	1005
Db	895	TTCTCTCTCTGTTGGTCTGGGCGACCAATGTTGTTGTCGGCATCCCGGTCACTCCGGC	954
Qy	1006	CAATCTACTTTCATGCTGGCCACCATGGTGATCGCGTGCAGCCGGCATTAAGATCTTC	1065
Db	955	GAGCTGTTCTTCATGTACGGCCACCATGCTGATCGCGTGCACCGGGTGAAGTGTTC	1014
Qy	1066	TCGTGGAATCCGACGATGTGGGCGGCTCGGTTGAGTTTCAAAATCGCGATGCTCTGGGCC	1125
Db	1015	AACGTGGTGACCAATGTGGGAGGTTCCGCTGACCTTCGAGACGCGATGCTGTTTCGCC	1074
Qy	1126	TTTGGCTTTATGTTTCGTTTACCGTGGGTGTGTGACCGGTATCGTGTGGGCCCAAGCG	1185
Db	1075	GTTGCCCTTGCTATCTCTGTTACCAATCGGGGCTTCTCGAGATGATGCTGGCGATCGCC	1134
Qy	1186	GGTCTGACCGTGTATATCAGCACCTATTACGTGTGGCGACTTCCATATATGTGATG	1245
Db	1135	CCGCGACAGTTCCAGTACCAGACACCTACTTCTGTGTCGCCCACTTCCACTACGTGCTG	1194
Qy	1246	TCGCTGGGTGCGATCTTTGGATCTTCGCGGTATCTACTTTTACATGCCGAAGTTCTCG	1305
Db	1195	GTGCGCGCGCGATCTTCGGCATCTTCGCTTCGCTTACTACTGGCTGCCGAAGTGGACC	1254
Qy	1306	GGCGCGCTTTTCCGCGAATGGGCTGCAAGGTGCACCTTCTGGACCTTCTTATCGTGGCG	1365
Db	1255	GGCCACATGTACGACGAGACCTTCGGCAAGCTGCACTTCTGGATGAGCTTCATCGGCATG	1314
Qy	1366	AAGTACAGTCTTTCGCGACACTTCTCTGGAGGTACAGGTATGCGCGCGGTATCATC	1425
Db	1315	AACCTGGCGTCTTCCCGATGCACTTCGTGCGCTTCGCGGCATGCGCGACGATCCCC	1374
Qy	1426	GACTATCCGAAAGCTTCGCGCTGTGAAACAAAGTCTCTGTCCTATGTTGGTTCCTGGCC	1485
Db	1375	GACTACACTGTGAGTTCGCGCACTTCAACATGTCTCTGTCGATCGCGCGCTTCATG---	1431
Qy	1486	TTGCGCTCGTTCCTGTTCTTCATCGTGAATCTTTGTATACGCTGTTGTCTGGCGCGCGC	1545
Db	1432	TTGCGCACCAACCAAGTCTGTTCTCTGTTTCATCGTTCATCAAGTGCATCCGGCGGGCAAG	1491
Qy	1546	GAGACCGTTCGMAACCGGTGGGCGAATTCGCCGATACGCTGGAAATGACGCTGCCATCA	1605
Db	1492	CGGGCCCCCTGCAACGCTTGGACG---GCGCGGAGGCGCTTGGATGGAGCATCCCCCTCG	1548
Qy	1606	CGGCTCCGGGCCACACGTTTGAACAGCTGCCCAAG	1641
Db	1549	CGGGGCGCTTACCAACCTTTCAGCACCCGCGCGAG	1584

```

/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELIUPA 034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 25546
/ LENGTH: 1629
/ TYPE: DNA
/ ORGANISM: Mycobacterium avium
/ US-10-282-122A-25546

```

RESULT 3
US-10-282-122A-25546
; Sequence 25546, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari

```
Db 564 CCTGACCGCGCGTGTTCGGGCTGGCGCGCGACCGGCATCTGGCGGCCCATGTCTACGA 623
Qy 768 TCCTGCTGGCGCGGTGACCGATTCTGTACCAACACATCTCTGGTTCCTTTGGGACCC 827
Db 624 CGCGGCCAACCGCGGAGTTC---TGTGTGGCAGCACCCTGTTCTGGTTCCTGGCCATCC 680
Qy 828 GGAAGTGTACATCATATTCGCGCGCTTTGGCGATCATGAGCATGCTGTGTCGACCTT 887
Db 681 CGAGGTGTACATCATCGGTTCGCGTCTTCGGGATCATCAACGAGATCATCCGCTGT 740
Qy 888 CTCGAAAAGCGCTTCCTGGTTCACCTGCGATGGTCTAGCAATGGTGCATCGTGT 947
Db 741 CGCGCGAAGCGCGGTGTTCGGCTACACACCGCTGGTGTAGCGACGCTGTCGATCCGGC 800
Qy 948 TCTGGGCTTTGCTGCTGGCGCGACCATGTATACACCGTTGGTATGTGCTGACCCAGCA 1007
Db 801 GCTGTCGCTCGCGTGTGGCGCGACCATGTTCCGCCACCGAGCGCTTCTGCTGCGGT 860
Qy 1008 ATCCTACTTATCTGCGCCACCATGTGTATCGCGTGGCGACCGGCATTAAGATCTTCTC 1067
Db 861 CTTTTCGTTTCATG-----ACGTACCTGTATCGCGGTGCCGACCGGATCAAGTCTTCAA 914
Qy 1068 GTGGATCGCCACGATGTGGCGCGCTCGGTGTAGTTCAAAATCGCGATGCTCTGGSCCTT 1127
Db 915 CTGGATCGGCACGATGTGGAAGGGGAGTTGACATTTGAGACCGCATGCTGTCTGCGT 974
Qy 1128 TGGCTTTATGTTCTGTTTCAACCTGGGTGTGTGACCGGTATCGTCTGGCCCAAGCGG 1187
Db 975 CGGCTTCTCTCACTTCTGCTGCTGGGTGTGTGACCGCGGTATGCTGCGCCAGCCGCC 1034
Qy 1188 TCTGGACGCTGATATCAGCACACCTATTACGCTGTGGCGACTTCCATATGTATGTC 1247
Db 1035 GCTGGACTTCCAGTACCGCACACTATTTCGTTGGTGGCGACTTCCATCTAGCTGCTGT 1094
Qy 1248 GCTGGGTGCGATCTTTCGCGATCTTCGCGGTATCTACTTTTACATCGGAGTTCCTCGG 1307
Db 1095 CGGCACCATCGTGTTCGCCACCTTCGCGGGGTCTACTTCTGTTCCCGAAGTACCGG 1154
Qy 1308 CCGCGCTTTCGCGAATGGGCTCAAGCTGCACTTCTGCACTTCTTATCGGTGCGNA 1367
Db 1155 CCGCTGCTCGACGAGCGGTGGGCAAGTGTCACTTCTGTTGACGTTTCACTGGGTCCA 1214
Qy 1368 CGTCAGCTTCTTCCGACGACTTCTTGGGACGTCAGGATGTCGCGCGCGTTCATCGA 1427
Db 1215 CACACCTTCTGCTGACGACTGCTGGCGGATTTGGGCACTTGGCGCGCTACGCGCA 1274
Qy 1428 CTATC-----CCGAGCCTTCGCGTGTGGAAACAAGTCTGTCCTATGTTGGTCTCT 1481
Db 1275 CTACCTGCCAGCGAGCGGTTCAGCCCTACACGTCGCTCCACGCTCGGCGCGTTCAT 1334
Qy 1482 GGCCTTCGCTCGTCTGCTTCTTCTTCTATCTGATCTTGTCTATACGCTGTTGCTGGCG 1541
Db 1335 CTTGGGCGCTCGATGTTCCGTTCTGCTTGGAACTCTTCAAGAGCTGCGCTACGCGCA 1394
Qy 1542 CCGCGAGACCGTTCGAAACCGTGGGCGAATTCGCCGATACGCTGGAATGACGCTGCC 1601
Db 1395 GGTGTCACCGTTCGAGACCGCGTGGG---TTACGGCACTCGCTGGAGTGGCGGACAG 1451
Qy 1602 ATCACCCTTCGCGGCCACACGTTTCGAAACGCTGCC 1638
Db 1452 CTGCGCGCGCGCGGCACAACTTACCGAGCTGCC 1488
```

RESULT 4

```
US-10-308-264-150
; Sequence 150, Application US/10308264
; Publication No. US20040029133A1
; GENERAL INFORMATION:
; APPLICANT: HerinStadt, Corinna
; TITLE OF INVENTION: MITOCHONDRIAL DNA POLYMORPHISM
; FILE REFERENCE: 660088.461
; CURRENT APPLICATION NUMBER: US/10/308,264
```

```
; CURRENT FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 697
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 150
; LENGTH: 16559
; <211> conflict
; TYPE: DNA
; ORGANISM: Homo sapiens mitochondrial
; FEATURE:
; LOCATION: 3106
; OTHER INFORMATION: "n" refers to a nucleotide at position 3107 of the
; OTHER INFORMATION: Cambridge Reference Sequence (CRS, Anderson et al.
; OTHER INFORMATION: 1981 Nature 290:457-465), later determined to be
; OTHER INFORMATION: absent from the human mtDNA sequence according to
; OTHER INFORMATION: the revised CRS (rCRS, Andrews et al. 1999 Nature
; OTHER INFORMATION: Genetics 23:147); "n" therefore represents a
; OTHER INFORMATION: single nucleotide deletion at a position corresponding to
; OTHER INFORMATION: nucleotide 3107 of SEQ ID NO:1 as provided herein.
US-10-308-264-150
```

```
Query Match 23.8%; Score 398.8; DB 16; Length 16559;
Best Local Similarity 58.2%; Pred. No. 5e-110;
Matches 744; Conservative 0; Mismatches 522; Indels 12; Gaps 2;

Qy 244 AACGACACCTGTGGAACGTCATGTTACCTACCATCGTATCTGATGATGTTCTTTGTG 303
Db 6050 AACGACACCATCTACAACGTTATCGTCACAGCCCATGATTTGTAATAATCTTTTCATA 6109
Qy 304 GGTATCCCGCATTTTCGGTGGTTTTGTAACATCTGATCGCGTGCAAAATCGGCGCT 363
Db 6110 GTATATCCCATCAATAACGGAGGCTTTGGCAACTGACTAGTTCCCTTAATAATCGGTGCC 6169
Qy 364 CCGGATATGGCTTCCCGCGTATGAACAACCTGCTGCTTCTGGCTGTTCATTTGCGGTACC 423
Db 6170 CCGGATATGGCTTCCCGCGCATAAACAATAAGCTTCTGACTTCTACCTCCCTCTCTC 6229
Qy 424 CGGATGGCGGTGCTTCGCTTTCGACCGCGGGTGCAGCTCAGCTGGGTTCGGGCGTT 483
Db 6230 CTACTCTGCTCGCATCTGTATAGTAGAGCGCGAGCAGAAACAGTTGAAACAGTC--- 6286
Qy 484 GGTGGGTTCGTACCCGCGCTGTGCAACCGGGAAGCTGGCTATTGATGAGACCTCGCG 543
Db 6287 -----TACCTCCCTTAGCAGGAACTACTCCACCTCGAGCTCCGTAGACTAAC 6340
Qy 544 ATTTTCGGGTTCACTTTCGGGTGCTCTCGATCATGGCGCGATCAACATGATCAG 603
Db 6341 ATCTTCTCTTACACCTAGCAGGTGCTCTCTATCTTAGGGGCCATCAATTTTCATACA 6400
Qy 604 ACCTTCTTGAACATGCGCGCCCGCATGACGTCGACAAAGTGGCGTGTTCGTGTTG 663
Db 6401 ACAATTATCAATAAAACCCCTGCCATAACCAATACCAACGCCCTTTCGTCTGA 6460
Qy 664 TCGATCTTATACAGGTGCTGCTGCTGCTGCGGTGCGGTCTGGCTGTGCAATC 723
Db 6461 TCGGCTCTAATACAGCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6520
Qy 724 ACATGCTGTCGACCGACCTAACTTCGGCACCACTTCTTCAATCTGCTGGCGCGGT 783
Db 6521 ACTATACTACTACAGACCGCAACCTCAACACACCTTCTTCGACCCCGCGGAGGGA 6580
Qy 784 GACCGGATTCGTACCAACATCTCTGTGGTCTTTGGGCAACCCGGAAGTGTACATC 843
Db 6581 GACCCCATTTCTATACCAACACCTATTTCTGATTTTTCGGTCACTGAAAGTTATTTCT 6640
Qy 844 ATTCTGCGCGCTTTGGCATCATCAGCATGCTGCTGCGACCTTCTC---GAAAGCGG 900
Db 6641 ATCTTACAGGCTTCGGAATAATCTCCATATTTGTAACCTACTCTCCGAAAAAAGAA 6700
Qy 901 GTCTCGGTTACCTGCGGATGTTCTATGCAATGGTGGCAATCGGTGTTCTGGGCTTTGTC 960
Db 6701 CCATTTGGATACATAGTATGTTCTGAGCTATGATCAATTAATTAATTAATTAATTAAT 6760
Qy 961 GTCTGGGCGCACCATGTACACCGTGGTATGTCGCTGACCCAGCAATCTTACTTCTATG 1020
```

6761 GTGTGAGCACACCATATATTTACAGTAGGAATAGACGTAGACACAGGACATATTTCCACC 6820
Db
1021 CTGGCCACCATGGTGTATCGGGTGGCGACCGGATTAAGATCTTCCTGGATCGCCAG 1080
Qy
6821 TCCGCTACCATATATCTGATATCCACCGGGGTCAAGATATTAGCTGACCTGCCACA 6880
Db
1081 ATGTGGGGGGCTCGGTGTAGTTCAAAATCCCGGATGCTGTGGCCCTTTGGCTTTATGTC 1140
Qy
6881 CTCCACGGAAGCAATATGAATGATCTGTGCGAGTGTCTGAGCCCTAGGATTCATCTTT 6940
Db
1141 CTGTTACCGGTGGGTGTGACCGGTATCGTCTGGCCCAAGCGGTCTGGACCGTGCA 1200
Qy
6941 CTTTTACCGGTAGTGGCTGACTGGCATTTGATTAGCAAACTCATCATAGACATCGTA 7000
Db
1201 TATCAGACACCTATTACGTGGTGGCGCACTTCATATATGTGATGCGTGGTGGCATC 1260
Qy
7001 CTACACGACAGTACTAGTGTAGCTCACTTCCACTATGCTCATATGCTCAATAGGAGCTGTA 7060
Db
1261 TTTCCGATCTTCGGCGGTACTACTTTTACATCCGAAGTTCCTCGGGCGCGCTTTCCCG 1320
Qy
7061 TTTGCCATCATAGAGGCTTCATCTACTGATTTCCCGTATTTCTAGGCTTACACCTAGAC 7120
Db
1321 GAATGGGCTGAAAGCTGCTCTGGAACCTTTCTCATCGGTGCGAACCTGCTCTTC 1380
Qy
7121 CAAACCTACGCAAAATCCATTTCCGTATCATATTCATGCGGTAATCTAACTTTCTTC 7180
Db
1381 CCGCAGCACTTCCTGGGACGCTCAGGATATCGCGCGGTTACATCGACTATCCGAGCC 1440
Qy
7181 CCACACACATTTCTCGGCTTATCCGGAATGCCCGGACGTTATCTCGGACTACCCGATGA 7240
Db
1441 TTCCGCTGTGGAAACAAAGTCTCGTCTATGTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
Qy
7241 TACACACATGAATATCTCATCTATCTGAGGCTCATCTATTTCTTAACAGCAGTAATA 7300
Db
1501 TTCTTCATCTGATCTTT 1518
Qy
7301 TTATAATTTTCATGAT 7318
Db

RESULT 5

US-10-308-264-145
; Sequence 145, Application US/10308264
; Publication No. US20040029133A1
; GENERAL INFORMATION:
; APPLICANT: HerinStadt, Corinna
; TITLE OF INVENTION: MITOCHONDRIAL DNA POLYMORPHISM
; FILE REFERENCE: 660088.461
; CURRENT APPLICATION NUMBER: US/10/308,264
; CURRENT FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 697
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 145
; LENGTH: 16561
; <211> conflict
; TYPE: DNA
; ORGANISM: Homo sapiens mitochondrial
; FEATURE:
; LOCATION: 3108
; OTHER INFORMATION: "n" refers to a nucleotide at position 3107 of the
; OTHER INFORMATION: Cambridge Reference Sequence (CRS, Anderson et al.
; OTHER INFORMATION: 1981 Nature 230:457-465), later determined to be
; OTHER INFORMATION: absent from the human mtDNA sequence according to
; OTHER INFORMATION: the revised CRS (rCRS, Andrews et al. 1999 Nature
; OTHER INFORMATION: Genetics 23:147); "n" therefore represents a
; OTHER INFORMATION: single nucleotide deletion at a position corresponding to
; OTHER INFORMATION: nucleotide 3107 of SEQ ID NO:1 as provided herein.
US-10-308-264-145

Query Match 23.8%; Score 398.8; DB 16; Length 16561;
Best Local Similarity 58.2%; Pred. No. 5e-110;
Matches 744; Conservative 0; Mismatches 522; Indels 12; Gaps 2;

244 AACGGACACCTGTGGAAACGTATGTTTACCTACCATGGTATTTCTGATGATGTTCTTTG 303
Qy
6052 AACGACACATCTACAAAGTTATCGTACAGCCATGATTTGTAATAATCTTTTCA 6111
Db
304 GGTATCCCGCATGTTGGTGTGTTTGTAACTATCTGATGCCCTGCAATCGCGCT 363
Qy
6112 GTAAATCCCATATAATCGGAGCTTTGGCACTGACTAGTTCCCTTAATAATCGGTGCC 6171
Db
364 CCGGATATGCTTCCCGGCTATGAACAACTGTCTGTTCTGGCTGTCTTCAATCCCGGTACC 423
Qy
6172 CCGGATATGCTTCCCGGCTATGAACAACTGTCTGTTCTGACTCTTACCTCCCTCTCTC 6231
Db
424 GCGATGGGCTGTGCTGCTGTTGCGACCGGGGGTACGGTACGCTGGGTTGGGCGTT 483
Qy
6232 CTACTCTGCTGCTGCTGCTGCTATAGTGGAGGCGGAGCAGGAACAGGTTGAACAGTCT 6288
Db
484 GGTGGGTTCTGTACCGCGCTGTGCGACCGCGAGCTGGCTATTCGATGACCTCGCG 543
Qy
6289 -----TACCTCTCCCTTAGCAGGAACTACTCCACCTGGAGCTCCGTAGACCTAAC 6342
Db
544 ATTTTCGGGTTTCACTGTGGGTGCTGCTGATGATGCGGCGCATCAACATGATCAG 603
Qy
6343 ATCTTCTCTTACACCTAGCAGTGTCTCTTATCTTAGGGGCCATCAACTCATACACA 6402
Db
604 ACCTTCTTGAACATGCGCGCGCGCGCATGAGCTGCACAAAGTGCCTGTTGTTCTCGTGG 663
Qy
6403 ACAATTATCAATATAAAACCCCTGCCATAACCCCAATACCAACGCCCTCTTCTGCTGA 6462
Db
664 TCGATCTTATCAGGCTTGGCTGATCTCTGGCGCTCCCGGTTCTGGCTGGTGAATC 723
Qy
6463 TCGCTCTTAATACAGCAGTCTTCTCTTCTCTCCAGTCTCTAGCTCTGGCATC 6522
Db
724 ACCATGCTGTGACCGACCGTAACTTTCGGCAGACCTTCTTCAATCTCTGCTGGCGCGCT 783
Qy
6523 ACTATACTACTACAGACCGCACTCAACACACCTTCTTCGACCCCGCGGAGGGA 6582
Db
784 GACCGATCTGTACCAACACATCTCTGTGTTCTTGGGCACCCCGGAAGTGTACATCATC 843
Qy
6583 GACCCATCTTACCAACACCTTATCTGATTTTCGGTCACTCCCTGAGTTATATTTCT 6642
Db
844 ATTCTGCCGCTTTCGGCATCATCAGCCATGCTGCTGACCTTCTC-----AAAAAGCG 900
Qy
6643 ATCTACAGGCTTCGGAATATCTCCCATATTTGTAATTTACTCTCCGGAATAAAGAA 6702
Db
901 GTCTTCGGTTACCTCGGATGTTCTATGCAATGGTGGCAATCGGTGTTCTGGGCTTTGTC 960
Qy
6703 CCAATTTGGATACATAGTATGCTGCTGAGCTATGATCAATTTGGGTTCTCTAGGTTTATC 6762
Db
961 GTCTGGCGCACCATGATACACCGTTGTTGCTGCTGACCCAGCAATCTTACTTCTATG 1020
Qy
6763 GTGTGAGCACCATATATTTACAGTAGGAATAGAGTAGACACAGGACATATTTTACC 6822
Db
1021 CTGGCCACCATGTTGATCGGGTGGCGGACCGGATTAAGATCTTCTGTTGATCGGACG 1080
Qy
6823 TCCGCTACCAATAATCATCTCTATCCCAACCGGCTCAAAAGTATTTAGCTGACTCGCCACA 6882
Db
1081 ATGTGGGGGGCTGCGGTTCAGTTCAAAATCGCGGATGCTCTGGGCTTTGGGCTTTATGTT 1140
Qy
6883 CTCACGGAAGCAATATGAATGATCTGCTGAGTCTGAGCCCTAGGATTCATCTT 6942
Db
1141 CTGTTACCGTGGTGTGTGACCGGTATCTGCTGGCCCAAGCGGCTTGGACCGTGA 1200
Qy
6943 CTTTTCACCGTAGGTGGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7002
Db
1201 TATCAGACACCTATTACCTGTTGGTGGCGCACTTCCATTTATGATGCTGCTGGGTCGGATC 1260
Qy
7003 CTACACGACAGTACTACCTGTTGAGTCTCACTTCCACTATGCTCTATCAATAGAGCTGTA 7062
Db
1261 TTTGGGATCTTCCCGGTATCTACTTTTACATGCGGAAGTCTCGGCGCGCGCTTTCCCG 1320
Qy
7063 TTTGGCATCATAGGAGCTTCAATCTGATTTCCCTTATTTCCAGGTACACCTAGAC 7122
Db
1321 GAATGGGCTGCAAAAGCTGCATCTTCGACCTTCTTTCATCGGTGCGAAGCTCACGTTCTTC 1380
Qy

Db 7123 CAACCTACGCGCAAAATCCATTTCACTATCATATATCATCGCGGTAAATCTAATCTTCTTC 7182
 QY 1381 CGCAGCACTTCTCGGGAGCTGACGGGTATGCGGGCGGTACATCGACTATCCCGAAGCC 1440
 Db 7183 CCACAACACTTCTCGGCTATCCGGAATGCCGAGCTTACTCGGACTACCCCGATGCA 7242
 QY 1441 TTCGGCTGTGGAACAAAGTCTCGTCTATGCTGGTTCCTGGCTTCGCTCGCTTCCTG 1500
 Db 7243 TACACCACATGAACATCCTATCATCTGTAGGCTCATTTCTCTAAACAGCAGTAATA 7302
 QY 1501 TTCCTCATCGTATCTTT 1518
 Db 7303 TTAATATTTTCATGATT 7320

RESULT 6

US-10-308-264-414
 ; Sequence 414, Application US/10308264
 ; Publication No. US20040029133A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HerinStadt, Corinna
 ; TITLE OF INVENTION: MITOCHONDRIAL DNA POLYMORPHISM
 ; FILE REFERENCE: 660088.461
 ; CURRENT APPLICATION NUMBER: US/10/308,264
 ; CURRENT FILING DATE: 2002-11-25
 ; NUMBER OF SEQ ID NOS: 697
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 414
 ; LENGTH: 16568
 ; <211> conflict
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens mitochondrial
 ; FEATURE:
 ; LOCATION: 3106
 ; OTHER INFORMATION: "n" refers to a nucleotide at position 3107 of the
 ; Cambridge Reference Sequence (CRS, Anderson et al.
 ; OTHER INFORMATION: 1981 Nature 290:457-465), later determined to be
 ; OTHER INFORMATION: absent from the human mtDNA sequence according to
 ; OTHER INFORMATION: the revised CRS (rCRS, Andrews et al. 1999 Nature
 ; OTHER INFORMATION: Genetics 23:147); "n" therefore represents a
 ; OTHER INFORMATION: single nucleotide deletion at a position corresponding to
 ; OTHER INFORMATION: nucleotide 3107 of Seq ID No:1 as provided herein.
 US-10-308-264-414

Query Match 23.8%; Score 398.8; DB 16; Length 16568;
 Best Local Similarity 58.2%; Pred. No. 5e-110;
 Matches 744; Conservative 0; Mismatches 522; Indels 12; Gaps 2;
 QY 244 AAGGACACCTGGAACGTCATGGTTACCTACATGGTATCTGATGATGTTCTTTG 303
 Db 6050 AAGGACACATCTACAACTGTTATCGTCACAGCCATGCTATTGTAATAATCTTCTATA 6109
 QY 304 GGTATCCCGCATGTTTCGGTGGTGTGTAATCTGATCGCTGCAATCGCGCT 363
 Db 6110 GTAATACCATCATATATCGGAGCTTGGCACTGACTAATTCCTTAAATATGCTGCT 6169
 QY 364 CCGGATATGCGCTCCCGGTATGAACAACTGTCTGCTGGCTGTTCATTCGCCGTACC 423
 Db 6170 CCGGATATGCGCTTCCCGCATGAACAACTGCTGCTGCTGCTGCTGCTGCTGCT 6229
 QY 424 GCGATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 483
 Db 6230 CTACTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6286
 QY 484 GGTGGGTTCTGTACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 543
 Db 6287 -----TACCCCTCCCTTAGCAGGAACCTACTCCACCCCTGGAGCCTCCGTAGACCTAAC 6340
 QY 544 ATTTTCGGGTTCACTTGTGCGGTGCTTCTGATCATGCGGCGGATCAACATGATCAG 603
 Db 6341 ATCTTCTCTACCTAGCAGGTGCTCTCTATCTTCTAGGGGCTCAATTTTCATCACA 6400

QY 604 ACCTTCTTGAAATGCGCGCCCGCATGACCTGCAAAAGTCCCGTGTGTTCTCGTGG 663
 Db 6401 ACAATATCAATATAAAACCCCTCGCATACCCAAATACCAAAACCCCTCTTCTGCTGA 6460
 QY 664 TGAATCTTTATCAGCGGTGGCTGATCTCTGCTGGCGCTGCGGTTTCGGCTGGTGAATC 723
 Db 6461 TCCGCTCTAATCAGACAGTCTTCTTCTCTATCTCTCCAGTCTTCTAGCTGCTGGCATC 6520
 QY 724 ACCATGCTGCTCAGCAGCGTAATCTTGGCAGACCTTCTCAATCTGCTGCGGCGGT 783
 Db 6521 ACTATACTACTAACAGACCGCACTCAACACCTTCTTCGACCCCGCGGAGGGA 6580
 QY 784 GACCGATCTGTACCAACATCTGTGGTCTTTGGGACCCCGGAAGTGTACATCATC 843
 Db 6581 GACCCCATTTCTATACCAACACTTCTGATTTTCGGTCACTGAGTTTATTTCTT 6640
 QY 844 ATTTCTGCGCGGTGGCATCATCAGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
 Db 6641 ATCTTACAGGCTTCGGAATATCTCCCATATTTGAACTTACTCTCGGAAAGAA 6700
 QY 901 GTCTTGGTACCTGCGCATGCTTATGCAATGGTGGCAATCGTGTCTGGGCTTTGTC 960
 Db 6701 CCATTTGGATACATAGGTATGCTGAGCTATGATATCAATTTGCTTCTAGGTTTATC 6760
 QY 961 GTCTGGGCGCACCATGATACACGTTGGTATGCTGCTGAGCCAGCAATCTTCTCATG 1020
 Db 6761 GTGTAGCAGCACCATATATTTTACAGTAGGAATAGACGTAGACACACGAGCATATTTCC 6820
 QY 1021 CTGGCCACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
 Db 6821 TCCGCTACCATATCATCTGCTATCCACCGCTCAAGTATTTAGCTGACTGCTGCTGCT 6880
 QY 1081 ATGTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
 Db 6881 CTCCAGCAAGCAATATGAAATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6940
 QY 1141 CTGTTACCGGTGGTGTGACCGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
 Db 6941 CTGTTACCGGTGGTGTGACCGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7000
 QY 1201 TATCAGCAGCACTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
 Db 7001 CTACAGCAGCACTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7060
 QY 1261 TTTGCGATCTGCGCGGTATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320
 Db 7061 TTTGCGATCTGCGCGGTATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7120
 QY 1321 GAATGGGCTGCAAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
 Db 7121 CAACCTTACGCCAAATTCATTTCACTATCATATTCATCGCGTAAATCTAATCTTCTTCT 7180
 QY 1381 CCGGAGCACTTCTGCGGAGCTGAGGTATCGCGGCGGTGTACATCGACTATCCCGAAGCC 1440
 Db 7181 CCACCACTTCTTCCGCGCTATCCGGAATGCCCGGAGCTTACTCGGACTACCCCGATGA 7240
 QY 1441 TTTGCGGTGTGGAACAAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
 Db 7241 TACACCATGAAACATCTTATCATCTGAGGCTCATTTCTTCTTAACAGCAGTAATA 7300
 QY 1501 TTTCTCATCTGATCTTT 1518
 Db 7301 TTAATAATTTTCATGATT 7318

RESULT 7

US-10-308-264-181
 ; Sequence 181, Application US/10308264
 ; Publication No. US20040029133A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HerinStadt, Corinna
 ; TITLE OF INVENTION: MITOCHONDRIAL DNA POLYMORPHISM
 ; FILE REFERENCE: 660088.461

```

; CURRENT APPLICATION NUMBER: US/10/308,264
; CURRENT FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 697
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 181
; LENGTH: 16569
; <211> conflict
; TYPE: DNA
; ORGANISM: Homo sapiens mitochondrial
; FEATURE:
; LOCATION: 3107
; OTHER INFORMATION: "n" refers to a nucleotide at position 3107 of the
; OTHER INFORMATION: Cambridge Reference Sequence (CRS, Anderson et al.
; OTHER INFORMATION: 1981 Nature 290:457-465), later determined to be
; OTHER INFORMATION: absent from the human mtDNA sequence according to
; OTHER INFORMATION: the revised CRS (rCRS, Andrews et al. 1999 Nature
; OTHER INFORMATION: Genetics 23:147); "n" therefore represents a
; OTHER INFORMATION: single nucleotide deletion at a position corresponding to
; OTHER INFORMATION: nucleotide 3107 of SEQ ID NO:1 as provided herein.
US-10-308-264-181

Query Match      23.8%; Score 398.8; DB 16; Length 16569;
Best Local Similarity 58.2%; Pred. No. 5e-110;
Matches 744; Conservative 0; Mismatches 522; Indels 12; Gaps 2;

QY 244 AACGACACCTGTGAACGCTGATGTTACCTACCATGGTATTCTGATGATGTTCTTGTG 303
DB 5051 AACGACACCATCTAACAGTTATGTCACAGCCGATGCAATTTGTAATCTTCTTCA 6110
QY 304 GGTATCCCGCGATGTTGCGTGTGTTGGTAATACTATCTGATGCGCTGCAATCGCGCT 363
DB 6111 GTAATACCCATCATATACGAGGCTTGGCAACTGACTAGTTCCCTTAATATCGGTGCC 6170
QY 364 CCGGATATGCGCTTCCCGGATGACACCTGTCGTTGCTGCTTCAATTCGCGGTACC 423
DB 6171 CCGGATATGCGGTTTCCCGGATGACACCAATGAGCTTCTGACTTCTACCTCCCTCTC 6230
QY 424 GCGATGGGCGTGGCTTTCGCTGTTCGACCGCGGTGACGGTCAGCTGCGGTTCGGCGGT 483
DB 6231 CTACTCTCTCGCATCTCTATATGAGGAGCGGAGCAGGAGGTTGAACATGCT--- 6287
QY 484 GGTGGGTTCTGTACCCCGCGTGTGACCCCGGAGCTGGCTATTGATGACCTCGG 543
DB 6288 -----TACCCTCCCTTAGCAGGAACTACTCCACCTCGGAGCGCTCGTAGACCTAAC 6341
QY 544 ATTTTCGCGGTTTCACTTGTGCGGTGCTCTCGATCATGGCGCGATCAACATGATCAG 603
DB 6342 ATCTTCTCTTACCTAGAGGTGCTCTCTATCTTAGGGCCATCAATTTCAATCACA 6401
QY 604 ACCTTCTTGAACATGCGCGCCCGCGCATGACGTGCAAAAGTGCGGTGTTCTCGTGG 663
DB 6402 ACAATATCAATATAAAACCCCTGCGCATACCCCAATACCAACGCCCTCTTCTCTGA 6461
QY 664 TCGATCTTTATACGCGTTGGCTGATCCTGCTGGCGCTCGCGTCTGCTGCTGCTGCAATC 723
DB 6462 TCGCTCTTAATACAGAGTCTTCTCTCTATCTCTCCAGTCTAGCTGCTGCTGATC 6521
QY 724 ACCATGCTGTGACCGACCGTAACTTCGGCAGACCTTCTTCAATCTGCTGCGCGCGT 783
DB 6522 ACCATACTACTACAGACCGCAACCTCAACACACCTTCTTCGACCCCGCGGAGGAGGA 6581
QY 784 GACCGATCTGTACCAACATCTCTGTTGGTCTTGGCACCAGGAGTGTACATCATC 843
DB 6582 GACCCCATCTTATACCAACACCTTCTGTATTTTCGGTCAACCTTCTGATTTTCTTAT 6641
QY 844 ATTCTGCGCGGCTTGGCATCATCAGCCATGTCGTCGACCTTCTC---GAAAGCGCG 900
DB 6642 ATCTTACGAGCTTCGGAATATCTCCCATATTGTAACCTACTACTCCGGAAGAAAGAA 6701
QY 901 GTCTTCGTTACTGCGGATGTTGATGCAATGGTGGCAATGGTGTCTTGGCTTTGTC 960
DB 6702 CCATTGGATACATAGTATGGTCTGAGCTATGATATCAATTTGGCTTCTAGGTTTATC 6761
```

```

961 GTCTGGCGGACCAACATGTAACCGTTGGTATGTCGCTGACCCAGCAATCTTACTTCATG 1020
DB 6762 GTGTGAGCACACCATATATTTACAGTAGGAATAGACGTAGACACAGCATATTTTACC 6821
QY 1021 CTGGCCACCATGTTGATCGGGTGGCGACCGGCATTAAGATCTTCTCGTGGATCGCCAG 1080
DB 6822 TCGCTACCATATATCATCGCTATCCCGACCGCGTCAAGTATTTAGCTGACTCGCCACA 6881
QY 1081 ATGTGGGGCGGCTCGGTTGATTTCAATCCCGATGCTCTGGGCTTTGGCTTTATGTTTC 1140
DB 6882 CTCACGGAAGCAATATGAAATGATCTGCTGCAAGTCTCTGAGCCCTTAGGATTCATCTTT 6941
QY 1141 CTGTTTACCGTGGTGTGTGACCGGTATCGTCTGCGCCCAAGCGGTCTTGGACCGTGCA 1200
DB 6942 CTTTTACCGTAGTGGCTGACTGGCATTTGATTAGCAAACTCATCACTAGACATCGTA 7001
QY 1201 TATCAGACACCTATTACGTGGTGGCGCATCTCCATTTATGATGTCGTGGGTGGGATC 1260
DB 7002 CTACACGACACGTACTACGTTGTAGCTCACTTCCACTATGCTCTATCAATAGGAGCTGA 7061
QY 1261 TTTGCGATCTTCGCGGTATCTACTTTTACATCGGAAGTTCTCGGCGCGGTTTCCCG 1320
DB 7062 TTTGCCATCATAGAGGCTTCATCTCACTGATTTCCCTTATCTCAGGCTACACCTAGAC 7121
QY 1321 GAATGGGCTGCAAGCTGCACTTTCTGGACCTTTCTTCATCGTGGCGGATCGACGTCTTC 1380
DB 7122 CAACCTTACGCCAAATATCCATTTCACTATCATATTCATCGCGGTAAATCTAACTTTCTTC 7181
QY 1381 CCGGACGACTTCTCGGACGCTCAGGATGTCGGCGGTTACATCGACTATCCGAGGCC 1440
DB 7182 CCACACACCTTTCTCGGCTTATCCGGAATGCCCGAGTTACTCGGACTACCCCGATGCA 7241
QY 1441 TTCCGCTGTGGAAACAAAGTCTGCTCTATGTCGCTTCTCGGCTTTCGCTTCTGCTG 1500
DB 7242 TACACACATGAACATCTCTATCATCTGATGCTCAATTTCTCTACAGCAGTAATA 7301
QY 1501 TTCTTCATCGTATCTTT 1518
DB 7302 TTAATAATTTTTCATGATT 7319

RESULT 8
US-10-308-264-440
; Sequence 440, Application US/10308264
; Publication No. US20040029133A1
; GENERAL INFORMATION:
; APPLICANT: HerinStadt, Corinna
; TITLE OF INVENTION: MITOCHONDRIAL DNA POLYMORPHISM
; FILE REFERENCE: 650088.451
; CURRENT APPLICATION NUMBER: US/10/308,264
; CURRENT FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 697
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 440
; LENGTH: 16569
; <211> conflict
; TYPE: DNA
; ORGANISM: Homo sapiens mitochondrial
; FEATURE:
; LOCATION: 3107
; OTHER INFORMATION: "n" refers to a nucleotide at position 3107 of the
; OTHER INFORMATION: Cambridge Reference Sequence (CRS, Anderson et al.
; OTHER INFORMATION: 1981 Nature 290:457-465), later determined to be
; OTHER INFORMATION: absent from the human mtDNA sequence according to
; OTHER INFORMATION: the revised CRS (rCRS, Andrews et al. 1999 Nature
; OTHER INFORMATION: Genetics 23:147); "n" therefore represents a
; OTHER INFORMATION: single nucleotide deletion at a position corresponding to
; OTHER INFORMATION: nucleotide 3107 of SEQ ID NO:1 as provided herein.
US-10-308-264-440

Query Match      23.8%; Score 398.8; DB 16; Length 16569;
Best Local Similarity 58.2%; Pred. No. 5e-110;
Matches 744; Conservative 0; Mismatches 522; Indels 12; Gaps 2;
```


QY 244 AACGACACCTGTGAAAGCTCATGTTACCTACCATGTTATTCGTGATGATGTTCTTTGTG 303
Db 6051 AACGACACACATCTCAAGCTTATCGTCACAGCCCATGCAATTTGTAATATCTTTTCATA 6110
QY 304 GGTATCCCGCATTTGTCGGTGGTTTGGTAACATATCTGATCGCGCTGCAATTCGGCGCT 363
Db 6111 GTAATACCACATCAATTCGGAGGCTTTGGCAACTGACTAGTTCCTCCCTAAATTCGGTGCC 6170
QY 364 CCGGATATGGCTTTCCCGCGTATGAACAAACCTGTCGTTCTGGCTGTTTCATTTGCGCGGTACC 423
Db 6171 CCCGATATGGCTTTCCCGCGATAAACACATAGCTTTGACTCTTACCTCCCTCTCTC 6230
QY 424 CGATGGCGTGGCTTCGCTGTCGACCGCGGCGGTGACGTCAGCTGGGTTCGGGGT 483
Db 6231 CTACTCTGCTCGCATCTGCTATAGTGGAGCCCGGACGAAACAGGTTGAACAGTC--- 6287
QY 484 GTTTGGGTTCTGTACCGCGCTCTCGACCGCGAAGCTGCTATTTCGATGGACCTCGCG 543
Db 6288 -----TACCTCTCTTAGCAGGAACTACTCCACCCCTGGAGCCTCGTAGACCTAAC 6341
QY 544 ATTTTCGGGTTCATCTGTGGGTGCTCTCTCGATCATGGCGCGATCAACATGATCAAG 603
Db 6342 ATCTTCTCTTACACCTAGCAGGTGTCTCTCTATCTTAGGGGCACTCAATTTTCATCACA 6401
QY 604 ACCTTCTTGACATGCGCGCCCGCGCATGACGCTGCACAAAGTGGTGTCTCGTG 663
Db 6402 ACAATTAATCAATAAACCCTCTGCCATAACCCATACCAAGCCCTCTTTCGCTGA 6461
QY 664 TCGATCTTTATCAGCGTTGCTGATCTCTGCTGCGCTGCGGTTCTGGCTGGTGAATC 723
Db 6462 TCCGTCCTAATCAGCAGCTACTTCTCTCTCTCTCCAGCTCTAGCTGCTGCATC 6521
QY 724 ACCATGCTGTGACGACGCTAATCTCGGCACGACCTTCTCAATCTGCTGCGCGGT 783
Db 6522 ACTATATCTAAGACAGCCGCACTCAACACCACTTCTTCGACCCCGCGGAGGAGA 6581
QY 784 GACCCGATCTGTACCAACACATCTCTGCTGTTCTTGGGACCCGGAAGTATCATCATC 843
Db 6582 GACCCCATCTATACCAACACCTATCTGATTTTCGTCACCTGNAATTTATATCTT 6641
QY 844 ATTCGCGCGCTTTGGCATCAICAGCATCTGCTGTCGACCTTCTC-----GAAAAAGCCG 900
Db 6642 ATCTTACCGCTTCGGAATAATCTCCATATTGTAACTTACTACTCCGGAATAAGAA 6701
QY 901 GTCTTCGTTACCTGCGATGGTCTATGCAATGGTGGCAATGCTGTTCTGGGCTTTGTC 960
Db 6702 CCATTTGGATACATAGTATGGTCTGAGCTATGATATCAATTTGGCTTCTAGGTTTATC 6761
QY 961 GTCTGGGCGACCAATGTACCGTTGGTATGTCGCTGACCCAGCAATCCTACTTCAATG 1020
Db 6762 GTGTGAGCACCAATATATTACAGTAGGAATAGACGTAGACACAGGACATATTTCAC 6821
QY 1021 CTGCGCAACATGGTATCGGGTCCGACCGGATTAAGATCTTCTCGGATCGCCAG 1080
Db 6822 TCCGCTACCATATCATCGCTATCCCGACCGGCTCAAGATTTAGCTGACTCGCCACA 6881
QY 1081 ATGTGGGCGGCTCGGTGAGTTCAATCGCCGATGCTCTGGGCTTTGCTTTATGTTTC 1140
Db 6882 CTCACGGAAGCAATATGAATGATCTGCTGAGTGTCTGAGCCCTAGGATTTATCTTT 6941
QY 1141 CTGTTACCGTGGTGTGACCGGTATGCTGCTGGCCCAAGCGGTCTGACCCGTGA 1200
Db 6942 CTTTTCACCGTAGTGGCTGACTGGCAATGTAATTAGCAAACTCATCATAGACATCGTA 7001
QY 1201 TATCAGACACCTATTACGTGGTGGCGCACTTCCATTTATGATGTGCTGGGTGGGATC 1260
Db 7002 CTACACGACAGCTACTAGTTGTAGCTCACTTCCACTATGCTCTATCAATAGAGCTGA 7061
QY 1261 TTTGCGATCTTCCGCGTATCTACTTTTACATGCGGAAGTTCTCGGGCGCGCTTTCCCG 1320
Db 7062 TTTGCCATCATAGAGGCTTCATTCACTGATTTCCCTATTTCTCAGGCTACACCCCTAGAC 7121

RESULT 9

US-10-308-264-469
Sequence 469, Application US/10308264
Publication No. US20040029133A1
GENERAL INFORMATION:
APPLICANT: Herrnstadt, Corinna
TITLE OF INVENTION: MITOCHONDRIAL DNA POLYMORPHISM
FILE REFERENCE: 660088.461
CURRENT APPLICATION NUMBER: US/10/308.264
CURRENT FILING DATE: 2002-11-25
NUMBER OF SEQ ID NOS: 697
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 469
LENGTH: 16569
TYPE: DNA
ORGANISM: Homo sapiens mitochondrial
FEATURE:
LOCATION: 3107
OTHER INFORMATION: "n" refers to a nucleotide at position 3107 of the
OTHER INFORMATION: Cambridge Reference Sequence (CRS, Anderson et al.
OTHER INFORMATION: 1981 Nature 290:457-465), later determined to be
OTHER INFORMATION: absent from the human mtDNA sequence according to
OTHER INFORMATION: the revised CRS (rCRS, Andrews et al. 1999 Nature
OTHER INFORMATION: Genetics 23:147); "n" therefore represents a
OTHER INFORMATION: single nucleotide deletion at a position corresponding to
OTHER INFORMATION: nucleotide 3107 of SEQ ID NO:1 as provided herein.
US-10-308-264-469

Query Match 23.8%; Score 398.8; DB 16; Length 16569;
Best Local Similarity 58.2%; Pred. No. 5e-110;
Matches 744; Conservative 0; Mismatches 522; Indels 12; Gaps 2;
QY 244 AACGACACCTGTGGAAGCTCATGTTACCTACCATGTTATTCGTGATGATGTTCTTTGTG 303
Db 6051 AACGACACATCTCAACGTTATCGTCACAGCCCATGCAATTTGTAATATCTTTTCATA 6110
QY 304 GGTATCCCGCATTTGTCGGTGGTTTGGTAACATATCTGATCGCGCTGCAATTCGGCGCT 363
Db 6111 GTAATACCACATCAATTCGGAGGCTTTGGCAACTGACTAGTTCCTCCCTAAATTCGGTGCC 6170
QY 364 CCGGATATGGCTTTCCCGCGTATGAACAAACCTGTCGTTCTGGCTGTTTCATTTGCGCGTACC 423
Db 6171 CCCGATATGGGTTTCCCGCGATAAACACATAGCTTCTGACTCTTACCTCCCTCTCTC 6230
QY 424 CGATGGCGTGGCTTCGCTGTCGACCGGGCGGTGACGTCAGCTGGGTTCGGGGT 483
Db 6231 CTACTCTCTGCTCGCATCTGCTATAGTGGAGCCCGGACGAAACAGGTTGAACAGTC--- 6287
QY 484 GTTTGGGTTCTGTACCGCGCTCTGCGACCGCGAAGCTGCTATTTCGATGGAGCCTCGCG 543
Db 6288 -----TACCTCTCTTAGCAGGAACTACTCCACCCCTGGAGCCTCGTAGACCTAAC 6341
QY 544 ATTTTCGGGTTCATCTGTGGGTGCTCTCTCGATCATGGCGGATCAACATGATCAAG 603
Db 6342 ATCTTCTCTTACACCTAGCAGGTGTCTCTCTATCTTAGGGCCATCAATTTTCATCACA 6401

QY 961 GTCTGGGCGACCAATGTACACGGTTGGTATGTGCTGACCCAGCAATCTACTTCATG 1020
Db 6762 GTGTGAGCACACCAATATTTACAGTAGGAATAGACGTAGACACAGCATATTTTCAAC 6821
QY 1021 CTGGCCACCATGGTATCGGGTCCGACCGGCAATTAAGATCTTCTCGTGATCGGCAG 1080
Db 6822 TCCGCTACCAATATCATCGGTATCCGACCGGGTCAAGATATTTAGCTAGCTCGGCACA 6881
QY 1081 ATGTGGGCGCGCTCGGTGAGTTCAAAATCGCGATGCTCTGGGCGTTTGGCTTTATGTTT 1140
Db 6882 CTCACGGAAGCAATATGAATGATCTGCTGAGCCCTAGGATCACTTT 6941
QY 1141 CTGTTACCGTGGTGTGACCGGTATCGTGTGCGCCCAAGGGGTCTGGACCTGCA 1200
Db 6942 CTTTTCACCGTAGTGGCGCTGACTGGCAATTTGATTAGCAAACTCATCACTAGACATCGTA 7001
QY 1201 TATCAGCACACCTATTACGTGGTGGCGCACTTCCATTATGTGATGCTCGTGGGTGGATC 1260
Db 7002 CTACAGCACAGTACTAGTTGAGCCACTTCCACTATGCTTCTATCAATAGAGCTGTA 7061
QY 1261 TTTGCGATCTTCGCGGTATCTACTTTTATCATGCCGAAGTTCTCGGGCCCGCTTTCCCG 1320
Db 7062 TTTGCCATCATAGAGGCTTCATTCACTGATTTTCCCTATTTCTCAGGCTACACCCCTAGAC 7121
QY 1321 GAATGGGCTGCAAGCTGCACTTCTGACCTTCTTCACTGCGTGGGCGGACGTCAGTTCTTC 1380
Db 7122 CAACCTAGCCAAATCCATTTCACTATCATATTCATGCGGTAAATCTTAACTTTCTTC 7181
QY 1381 CCGCAGCACTTCTGGGACGTCAGGATGCGCGCGCTTACATCGACTATCCGGAAGCC 1440
Db 7182 CCAACACATTTCTCGGCTATCCGGAATGCGCGAGCTTACTCGGACTACCCCGATGCA 7241
QY 1441 TTGCGGTGTGGAAGAAGTCTGTCATGTTGCGGTCTTGGCCCTTCGCTCGTTCCTG 1500
Db 7242 TACACCAATGAACATCCCTATCATCTGATGCTAGGCTCATTTCTCTAACAGCAGTAATA 7301
QY 1501 TTTCTCATCTGATCTTT 1518
Db 7302 TTAATAATTTTCATGATT 7319

RESULT 11

US-10-308-264-476
; Sequence 476, Application US/10308264
; Publication No. US20040029133A1
; GENERAL INFORMATION:
; APPLICANT: Hermsstadt, Corinna
; TITLE OF INVENTION: MITOCHONDRIAL DNA POLYMORPHISM
; FILE REFERENCE: 660088.461
; CURRENT APPLICATION NUMBER: US/10/308,264
; CURRENT FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 697
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 476
; LENGTH: 16569
; TYPE: DNA
; ORGANISM: Homo sapiens mitochondrial
; FEATURE:
; LOCATION: 3107
; OTHER INFORMATION: "n" refers to a nucleotide at position 3107 of the
; OTHER INFORMATION: Cambridge Reference Sequence (CRS, Anderson et al.
; OTHER INFORMATION: 1981 Nature 290:457-465), later determined to be
; OTHER INFORMATION: absent from the human mtDNA sequence according to
; OTHER INFORMATION: the revised CRS (rCRS, Andrews et al. 1999 Nature
; OTHER INFORMATION: Genetics 23:147); "n" therefore represents a
; OTHER INFORMATION: single nucleotide deletion at a position corresponding to
; OTHER INFORMATION: nucleotide 3107 of SEQ ID NO:1 as provided herein.
US-10-308-264-476

Query Match

Best Local Similarity 23.8%; Score 398.8; DB 16; Length 16569;
Best Local Similarity 58.2%; Pred. No. 5e-110;

Matches 744; Conservative 0; Mismatches 522; Indels 12; Gaps 2;
QY 244 AACGACACCTGTGGACGTCATGTTTACCTTACCATGTTATTTCTGATGATGTTCTTTGTG 303
Db 6051 AACGACACACATCTACAAGTTTATCGTCACAGCCCATGCTTGTAAATATCTTCTTATA 6110
QY 304 GGTATCCCGCATGTTTTCGGTGGTTTGGTAACTATCTGATGCGCGCTGCATAATCGGCT 363
Db 6111 GTATACCCATCATATAATCGAGGCTTTGGCAACTGACTAGTTCCCTTAATAATCGGTGCC 6170
QY 364 CCGGATATGGCTTCCCGCGTATGAACAACTGCTGCTTCTGGCTGTTCAATGCGCGTACC 423
Db 6171 CCGGATATGGGTTTCCCGCATTAACAATAAGCTTCTGACTTACCTTCCCTCTCTC 6230
QY 424 GCGATGGCGGTGGTTCCTGCTTTCGACCGGGGGGTGACGTCAGTGGTTCGGGGT 483
Db 6231 CTACTCTCTGCTCGCATCTGCTATATGAGGAGCGGAGCAGGAACAGGTTGAACAGTC --- 6287
QY 484 GGTGGGTTCGTACCCCGCTGTCGACCCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 543
Db 6288 -----TACCTCTCTTACGAGGGAATCTATCCACCTCGAGGCTCCGTAGACTAAC 6341
QY 544 ATTTTCGGGTTCATGTCGGGTGCTCTCTGATCATGTCGCGCGATCAACATGATCAAG 603
Db 6342 ATCTTCTCTTACACCTAGCAGGTGCTCTCTATCTTAGGGGCGCATCAATTTTCATACA 6401
QY 604 ACCTTCTTGAACATCGCGCGCCCGGATGACGTCGACAAAGTGCCTGTTGTTCTCGTG 663
Db 6402 ACAATATCAATATAAACCCTCGCCATAACCAATACCAACGCGCTCTTCTGTCGA 6461
QY 664 TCGATCTTTATCAGGCTTGGCTGATCTGCTGCGCTGCGGTTCTGGTGGTGGCAATC 723
Db 6462 TCGGCTCTTATCAGCAGTCTCTACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6521
QY 724 ACCATGCTGTCGACGACCTTAACTTCGGGACGCTTCTTCAATCTGCTGGCGGGGT 783
Db 6522 ACTATACTACTAACAGACCGCAACCTCAACACCACTTCTTCGACCCCGCGGAGGAGA 6581
QY 784 GACCCGATTCGTACCAACACATCTGCTGGTTCCTTGGGACCCCGGAAGTGTACATCATC 843
Db 6582 GACCCCATCTATACCAACACTATCTGATTTTTCGTCACCTCGAAGTTTATTTCTT 6641
QY 844 ATTCGCGCGCTTTGGCATCATCAGCATGCTGCTGCGACCTTCTC ---GAAAAAGCG 900
Db 6642 ATCTTACCGAGCTTCGGAATTAATCTCCCATATTTAACTTACTACTCCGGAAGAAAGAA 6701
QY 901 GTCTTCGGTTACCTGCGGATGCTATGCAATGTCGCAATCGGTGTTCTGGGCTTTGTC 960
Db 6702 CCAITTTGATACATAGGTATGTCGAGCTATGATATCAATTTGCTTCTCTAGGTTTATC 6761
QY 961 GTCTGGGCGCACCATGATACACCTTTGGTATGTCGTCACCCAGCAATCTACTTCATG 1020
Db 6762 GTGTGAGCACACCATATTTACAGTAGGAATAGACGTAGACACAGCATATTTTCAAC 6821
QY 1021 CTGGCCACCATGTCATCGCGTCCGACCGGCAATTAAGATCTTCTGTCGATGCGCACG 1080
Db 6822 TCCGCTACCAATATCATCGCTATCCCAACCGGCTCAAGTATTTAGCTGACTGCGCAC 6881
QY 1081 ATGTGGGCGGCTCGGTTGAGTTCAAAATCGCCGATGCTCTGGGCTTTGGCTTTATGTT 1140
Db 6882 CTCACGGAAGCAATATGAATGATCTGCTGAGCTGCTGAGCCCTAGGATTCATCTTT 6941
QY 1141 CTGTTACCGTGGTGGTGGTACCGGTATCGTGTGCGCCCAAGGGGTCTGGACCGTGA 1200
Db 6942 CTTTTCACCGTAGTGGCGCTGACTGGCATTTGATTAGCAAACTCATCACTAGACATCGTA 7001
QY 1201 TATCAGCACACCTATTACGTCGGTGGCGCACTTCCATTATGTCGCTGGGTGGCATC 1260
Db 7002 CTACAGCACAGTACTAGTTGAGCCCACTTCCACTATGCTTCTATCAATAGGAGCTGTA 7061
QY 1261 TTTGGATCTTCGCGGTATCTCTTTTACATGCGGAAGTCTCGGCGCGGCTTTCCCG 1320
Db 7062 TTTGCCATCATAGAGGCTTCATTCACTGATTTTCCCTATTTCTCAGGCTACACCCCTAGAC 7121

QY 1321 GAATGGGCTGCAAGCTGCACTTCTGGACCTTCTTTCATCGGTGCAAGCTGCAAGCTTCTTC 1380
Db 7122 CAAACCTACGCCCAAAATCCATTTCATCAATATTCATCGCGGTAAATCTAACTTCTTC 7181
QY 1381 CGCAGCAGCTTCTCGGAGCTGAGGTATGCGCGCGCTTATCATGCAATATCCCAAGCC 1440
Db 7182 CCACAACACTTCTCGGCTATCCGGAATGCCCGACGTTACTCGGAGCTACCCCGATGCA 7241
QY 1441 TTCGGCTGTGGAACAAAGCTCTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
Db 7242 TACACACATGAACATCTATCACTGTAGCTCATTCATTTCTTAACAGCAGTAATA 7301
QY 1501 TTCTTCATCGTGAATCTTT 1518
Db 7302 TTAATAATTTTCATGATT 7319

RESULT 12
US-10-308-264-501
; Sequence 501, Application US/10308264
; Publication No. US20040029133A1
; GENERAL INFORMATION:
; APPLICANT: Herrin, Corinna
; TITLE OF INVENTION: MITOCHONDRIAL DNA POLYMORPHISM
; FILE REFERENCE: 660088.461
; CURRENT APPLICATION NUMBER: US/10/308,264
; NUMBER OF SEQ ID NOS: 697
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 501
; LENGTH: 16569
; <211> conflict
; TYPE: DNA
; ORGANISM: Homo sapiens mitochondrial
; FEATURE:
; LOCATION: 3107
; OTHER INFORMATION: "n" refers to a nucleotide at position 3107 of the
; OTHER INFORMATION: Cambridge Reference Sequence (CRS, Anderson et al.
; OTHER INFORMATION: 1981 Nature 290:457-465), later determined to be
; OTHER INFORMATION: absent from the human mtDNA sequence according to
; OTHER INFORMATION: the revised CRS (rCRS, Andrews et al. 1999 Nature
; OTHER INFORMATION: Genetics 23:147); "n" therefore represents a
; OTHER INFORMATION: single nucleotide deletion at a position corresponding to
; OTHER INFORMATION: nucleotide 3107 of SEQ ID NO:1 as provided herein.

US-10-308-264-501
Query Match 23.8%; Score 398.8; DB 16; Length 16569;
Best Local Similarity 58.2%; Pred. No. 5e-110;
Matches 744; Conservative 0; Mismatches 522; Indels 12; Gaps 2;
QY 244 AACGGACACTGTGGAACGTCATGGTTACCTACATGGTATCTGATGATGTTCTTTGTG 303
Db 6051 AACGACCACTACAAAGTTATCGTCACAGCCATGCAATTTGTAATAATCTTTCATA 5110
QY 304 GGTATCCCGCATGTTGCGGTGTTTGTGTAATCTATCTGATCCGCTGCAATTCGCGCT 363
Db 6111 GTAATACCATCATATAATCGAGGCTTTGGCACTGACTAGTTCCCTTAATAATCGGTGCC 6170
QY 364 CCGGATATGCTTCCCGGTATGAACACCTGCTGCTTGGGTGTTCAATTCGCGTACC 423
Db 6171 CCGGATATGCGTTTCCCGCATATAAACAATAGCTTCTGACTTCACTCCCTCTCTC 6230
QY 424 GCGATGGGCGTGGCTTGGCTGTTTCGACCGGCGGTGACGGTCAGGTGCGGTTCGGCGCTT 483
Db 6231 CTACTCTGCTGCACTGCTATAGTGGAGGCGGAGCAGGAACAGTTTGAACAGTC--- 6287
QY 484 GGTGGGTCTGTACCGGCGCTGTGACCCCGGAAGCTGGCTATTCGATGACCTCGCG 543
Db 6288 -----TACCCCTCTTAGCAGGAATACTCCACACCTCGGAGCCCTCCGTAGACCTAAC 6341
QY 544 ATTTTCGCGGTTCACCTGTCGGGTGCTCTCTGATCATGGCGCGATCAACATGATCAGC 603

RESULT 13
US-10-308-264-508
; Sequence 508, Application US/10308264
; Publication No. US20040029133A1
; GENERAL INFORMATION:
; APPLICANT: Herrin, Corinna

Db 6342 ATCTTCTCCTTACACCTAGCAGGTGTCCTCTATCTTAGGGCCATCAATTTTCATCACA 6401
QY 604 ACCTTCTTGAACATGCGCGCCCGCGCATGACGTGCAAAAGTGCCTGTTCTTCGTGG 663
Db 6402 ACAATATCAATATAAAACCCCTGCGCATAAACCAATACCAACGCCCTCTTCGTCTGA 6461
QY 664 TCGATCTTTATCAGCGTGTGCTGATCTCTGCGGTGCGGTGCTGCGGTGCTGCGGTGCTG 723
Db 6462 TCCGTCCTTAATACAGCAGTCTTCTCTCTATCTCTCCAGTCTAGTGTGCGCATC 6521
QY 724 ACCATGCTGTGACCGACCGTAACTTCGGCACACACCTTCTTCAATCTGCTGCGCGCGGT 783
Db 6522 ACTATCTACTACAGACCGCACTCAACACACCTTCTTCGACCCCGCGGAGGAGGA 6581
QY 784 GACCGGATCTGTACCAACACATCTCTGCTGTTCTTTGGGACCCCGGAAGTGTATCATC 843
Db 6582 GACCCCATTTCTATACCAACACCTATTCTGATTTTTCGCTCACCCTGAAGTTTATATCT 6641
QY 844 ATTCTGCCCGCTTTGGCATCATCAGCCATGTCGTGTCGACCTTCTC---GAAAGCCG 900
Db 6642 ATCTTACAGGCTTCGGAATAATCTCCATATTGTAACTTACTCTCGGAAAAAAGAA 6701
QY 901 GTCTTCGGTTTACCTGCCGATGGTCTATGCAATGGTGGCAATCGGTGTTCTGGGGTTGTC 960
Db 6702 CCAATTTGGATACATAGGTATGGTCTGAGCTATGATATCAATTTGGCTTCTTAGGGTTATC 6761
QY 961 GTCTGGGGGACCATGTCACCGTTGGTATGTCGCTGACCCAGCAATCTTACTTCTCATG 1020
Db 6762 GTGTGAGCACACCATATATTTTACAGTAGGAATAGACGTAGACACAGGAGATATTTTACC 6821
QY 1021 CTGGCCACCATGTCGCGGTGCCGCGGTGCCGCGGTAAAGATCTTCTCGTGAATCGCCACG 1080
Db 6822 TCCGCCACCAATATCATCGCTATCCCAACCGCGTCAAGTATTTAGCTGACTCGCCACA 6881
QY 1081 ATGTGGGGCGCTCGGTGATGCTCAAAATGCGCGATGCTCTGGGCTTGGCTTATGCTTC 1140
Db 6882 CTCACGGAAGCAATATGAATGATCTGCTGAGTCTGCTGAGCCCTAGGATTCATCTTT 6941
QY 1141 CTGTTACCGGTGGTGTGACCGGTATCGTGTGCGCCAAAGCGGTCTGGACCGTGA 1200
Db 6942 CTTTTCACGTAGTGGCTGACTGGCATGTTATAGCAAACTCATCACTAGACATGTA 7001
QY 1201 TATCAGACACTATATGCTGGTGGCGACTTCCATATGATGATGCTGCTGGGTGGGATC 1260
Db 7002 CTACACGACAGTACTACTGTTGAGCCACTTCCACTATGCTCTATCAATAGGAGCTGTA 7061
QY 1261 TTTGCGATCTTCCGCGGTATCTATTTTACATGCGGAAGTCTCGGCGCGCTTTCCTCG 1320
Db 7062 TTTGCCATCATAGAGGCTTCATTCATGATTTCCCTATTTCTCAGGCTACACCTAGAC 7121
QY 1321 GAATGGGCTGCAAGCTGCACTTCTGGACCTTCTTTCATCGGTGCGAAGCTGCACTTCTTC 1380
Db 7122 CAAACCTACGCCCAAAATCCATTTCACTATCATATTCATCGCGTAAATCTAACTTCTTC 7181
QY 1381 CCGGACACTTCTCGGAGCTGACGCTGATGCGCGCGGTATACATCGACTATCCGGAAGCC 1440
Db 7182 CCACAACACTTCTCGGCTATCCGGAATGCCCGACGTTACTCGGACTACCCCGATGCA 7241
QY 1441 TTCGCGCTGTGGAACAAAGTCTCGTCTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1500
Db 7242 TACACCATGAACATCTCTATCATCTGTAGGCTCATTTCTCTTAACAGCAGTAATA 7301
QY 1501 TTCTTCATCGTGAATCTTT 1518
Db 7302 TTAATAATTTTCATGATT 7319

; TITLE OF INVENTION: MITOCHONDRIAL DNA POLYMORPHISM

; FILE REFERENCE: 660088.461

; CURRENT APPLICATION NUMBER: US/10/308,264

; CURRENT FILING DATE: 2002-11-25

; NUMBER OF SEQ ID NOS: 697

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 508

; LENGTH: 16569

; <211> conflict

; TYPE: DNA

; ORGANISM: Homo sapiens mitochondrial

; FEATURE:

; LOCATION: 3107

; OTHER INFORMATION: "n" refers to a nucleotide at position 3107 of the

; OTHER INFORMATION: Cambridge Reference Sequence (CRS, Anderson et al.

; OTHER INFORMATION: 1981 Nature 290:457-465), later determined to be

; OTHER INFORMATION: absent from the human mtDNA sequence according to

; OTHER INFORMATION: the revised CRS (rCRS, Andrews et al. 1999 Nature

; OTHER INFORMATION: Genetics 23:147); "n" therefore represents a

; OTHER INFORMATION: single nucleotide deletion at a position corresponding to

; OTHER INFORMATION: nucleotide 3107 of SEQ ID NO:1 as provided herein.

US-10-308-264-508

Query Match 23.8%; Score 398.8; DB 16; Length 16569;

Best Local Similarity 58.2%; Pred. No. 5e-110;

Mismatches 744; Conservative 0; Mismatches 522; Indels 12; Gaps 2;

```
QY 244 AACGACACCTGTGGACGTCATGGTTACCTACCATGGTATCTCTGATGATGTTCTTTGG 303
DB 6051 AACGACACATCTACACGTTATGTCACAGCCCATGTCATTTGTAATAATCTTTTCA 6110
QY 304 GGTATCCCCGATGTTTCGGTGGTTTGGTAACTATCTGATGCGGTGCAAAATCGGCGT 363
DB 6111 GTAATACCATCAATAATCGGAGGCTTTGGCACTGACTAGTCCCTTAATAATCGTGCC 6170
QY 364 CCGGATATGGCTTCGGGATGAAACACTGTGTTCTGGTGTTCATATGCGGTACG 423
DB 6171 CCCGATATGGCTTCCCGCCATAAAACAACATAGCTTCTGACTCTTACCTCCCTCTC 6230
QY 424 CGGATGGGCGTGGCTTCGCTGTTCCGACCGGGGTGACGTCAGCTGGTTCGGGCGTT 483
DB 6231 CTACTCTGTCGCAATCTGCTATAGTGGAGCGGAGCAGGACAGTTGAACAGTC--- 5287
QY 484 GGTGGGTTCTGTACCCGCGCTGTGTGACCCGGAAGCTGGCTATTGATGGACCTCGCG 543
DB 6288 -----TACGCTCCCTTAGCAGGGAACACTCTCCACCCCTGGAGCCCTCGTAGACCTAAC 6341
QY 544 ATTTTCGGGTTCACTTGTGGGTGCTCTCGATCATGGCGCGGATCAACATGATCAG 603
DB 6342 ATCTTCTCTTACACCTAGCAGGTTCTCTCTATCTTAGGGCCATCAATTTTATCACA 6401
QY 604 ACCTTCTTGACATGCGCGCCCGGATGACGCTGCAAAAGTGGCTTTCTCGTG 663
DB 6402 ACAATTATCAATAFAAAACCCCTGCCATAACCAATACCAACGCCCTCTTCTGTGA 6461
QY 664 TCGATCTTTATCAGGCTTGGCTGATCTCTGCGGCTGCGGTTCTGGCTGTGCAATC 723
DB 6462 TCGGCTCTAATACAGCAGTCCCTACTTCTCTATCTCTCCAGTCTAGCTGTGGCATC 6521
QY 724 ACCATGCTGTGACCGACCGTAACTTCGGCAGCACCTTCTTCAATCTGCTGGCGCGGT 783
DB 6522 ACTATCTACTACAGACCGCACTTCAACACACCTTCTTTCAGCCCGCGGAGGAGA 6581
QY 784 GACCGGATCTGTACGAACATCTGTGTTCTTTGGGACCCGGAAGTGTACATCATC 843
DB 6582 GACCCCAATCTATACCAACACTATCTGATTTTTCGGTCAACCTGAAGTTATATCTT 6641
QY 844 ATTCGCGCGGCTTTGGCATCATCAGCATGTGCTGTGCACTTCTC---GAAAAAGCG 900
DB 6642 ATCTACAGGCTTCGAATAAATCTCCATATTTGTAACCTACTCTCCGGAATAAAGAA 6701
QY 901 GTCTTCGGTACCTGCGCATGGTCTATGCAATGGTGGCAATCGGTTCTTGGGCTTTGTC 960
```

```
DB 6702 CCAATTGGATACATAGGTATGGTCTGACGTATGATATCAATGGCTTCCTAGGTTTATC 6761
QY 961 GTCTGGGCGCACCATACATGATACACCGTTGGTATGCTGCTGACCCAGCAATCTACTTCATG 1020
DB 6762 GTGTGAGCACACCATATTTACAGTAGGAAATAGACGTAGACACACGAGCATATTTTACC 6821
QY 1021 CTGSCCAGCATGGTATCGCGTGGCCGACCGGATTAAGATCTTCTCGTGGATCGGCACG 1080
DB 6822 TCCGCTACCAATAATCATCGCTATCCCCACCGCGTCAAAAGTATTAGCTGACTCGCCACA 6881
QY 1081 ATGTGGGCGGCTCGGTTGAGTTCAAATCGCCGATGCTCTGGGCTTTGGGCTTTATGTT 1140
DB 6882 CTCACGGAAGCAATATGAATGATCTGCTGAGTGTCTGAGCCCTAGGATTCATCTTT 6941
QY 1141 CTGTTACCGTGGTGGTGTGACCGGTATCGTGTGCGCCAAAGCGGTCTGGACCGTGCA 1200
DB 6942 CTTTTCACCGTAGTGGCTGACTGGCAATTTGATTAGCAAACTCATCACTAGACATCGTA 7001
QY 1201 TATCAGACACCTATTAGTGGTGGCGCACTTCATTATGATGATGCTGCTGGGTGCGATC 1260
DB 7002 CTACAGACAGTACTAGTGTAGCCCACTTCCACTATGCTCTATCAATAGGAGCTGTA 7061
QY 1261 TTTCGATCTTTGCGCGGTATCTACTTTTACATGCGGAAGTTCTCGGCGCGGCTTTCCCG 1320
DB 7062 TTTCGATCTAGGAGGTTTCACTCACTGATTTCCTCCCTTATCTCAGGCTACACCTAGAC 7121
QY 1321 GAATGGGCTGCAAAAGCTCACTTCTGGACCTTCTTCGCGTGGCAACGTCACGTTCTTC 1380
DB 7122 CAACCTACGCCAAATCCATTTTCACTATATATCATGCGCGTAAATCTTAACCTTTCTTC 7181
QY 1381 CCGCAGCACTTCTTCGGGACGTGAGGTATGCGCGCGGTTTACATCGACTATCCCGAAGCC 1440
DB 7182 CCACAACACTTCTCGGCTATCCGGAATGCCCGAGTTACTCGGACTACCCCGATGCA 7241
QY 1441 TTGCGGCTGTGGAACAAGTCTGCTCTATGTTGGTTCCTGGGCTTCGCTGCTGCTGCTG 1500
DB 7242 TACACCATCAAAACATCTCTATCATCTGTAGGCTCACTTCATTTCTTAACAGCAGTAATA 7301
QY 1501 TTTCTTCATCGTGAATCTTT 1518
DB 7302 TTAATAATTTTCATGATT 7319
```

RESULT 14

US-10-308-264-524

; Sequence 524, Application US/10308264

; Publication No. US20040029133A1

; GENERAL INFORMATION:

; APPLICANT: Herznstadt, Corinna

; TITLE OF INVENTION: MITOCHONDRIAL DNA POLYMORPHISM

; FILE REFERENCE: 660088.461

; CURRENT APPLICATION NUMBER: US/10/308,264

; CURRENT FILING DATE: 2002-11-25

; NUMBER OF SEQ ID NOS: 697

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 524

; LENGTH: 16569

; <211> conflict

; TYPE: DNA

; ORGANISM: Homo sapiens mitochondrial

; FEATURE:

; LOCATION: 3107

; OTHER INFORMATION: "n" refers to a nucleotide at position 3107 of the

; OTHER INFORMATION: Cambridge Reference Sequence (CRS, Anderson et al.

; OTHER INFORMATION: 1981 Nature 290:457-465), later determined to be

; OTHER INFORMATION: absent from the human mtDNA sequence according to

; OTHER INFORMATION: the revised CRS (rCRS, Andrews et al. 1999 Nature

; OTHER INFORMATION: Genetics 23:147); "n" therefore represents a

; OTHER INFORMATION: single nucleotide deletion at a position corresponding to

; OTHER INFORMATION: nucleotide 3107 of SEQ ID NO:1 as provided herein.

US-10-308-264-524

Query Match

23.8%; Score 398.8; DB 16; Length 16569;

	Best Local Similarity	58.2%;	Pred. No. 5e-110;	
	Matches	744;	Conservative	0; Mismatches 522; Indels 12; Gaps 2;
QY	244	AACGGACACCTGTGGAAACGTCATGGTTACCTACATGGTATTCTCGATGATGTTCTTTGTG	303	
Db	6051	AACGACACATCTACAACGTTATCGTCACAGCCCATGCAATTTGTAATAATCTTCTTCATA	6110	
QY	304	GGTATCCCGCANTGTTCGGTGGTTTTGGTAACTATCTGATGCCGTGCAAAATCGGGCGT	363	
Db	6111	GTAATACCCATATAATTCGGAGCGTTTGGCAACTGACTAGTTCGCCATAATATCGGTGCC	6170	
QY	364	COGGATATGGCCTTCGGCGGTATGAACAACCTGTGTTCTGGTGTTCATTGCGCGTAACC	423	
Db	6171	CCCGATATGGCGTTTCCCGCATAAACAATAGCTTCTGACTCTTACCTCCCTCTCTC	6230	
QY	424	GGATGGGCGTGGCTTCGTGTTTTCGACCGGGCGGTGAACGTCAGCTGGGTTCGGGCGTT	483	
Db	6231	CTACTCTGTCGCATCTGCTATAGTGGAGCGCGGACAGAAACAGTTTGAACAGTCT--	6287	
QY	484	GGTTGGGTTCTGTAACCGCGCGCTGCGACCGCGGAAGCTGGCTATTTCGATGGACCTCGG	543	
Db	6288	-----TACCCTCCCTTAGCAGGAACTACTCCACCCCTGGAGCTCCGTAGACCTAAACC	6341	
QY	544	ATTTTCGGGTTCACTTGTTCGGTGGCTCTTCGATCATGGCGCGATCAACATGATCAACG	603	
Db	6342	ATCTCTCCCTTACACCTAGCAGGTGTCCTCTATCTTAGGGCCATCAATTTCAACACA	6401	
QY	604	ACCTTCTTGAACTGGCGCGCCCGGCATGACCTGCACAAAGTGCCTGTTCTCGTGG	663	
Db	6402	ACAATTATCAATATAAAACCCCTGCGMAAACCAATACCAAAACGGCCCTCTTCGTCTGA	6461	
QY	664	TCGATCTTTATACGGGTTGGCTGATCCTGCTGGCGCTGCGGTTCTGGCTGGTGGCAATC	723	
Db	6462	TCCGTCCTAATACAGCAGTCTTCTCTCTATCTCTCCAGTCCTAGCTGTGGCAAC	6521	
QY	724	ACCATGCTGTGACGACCGTAACCTTCGGACAGACTTCTTAAATCTCTGTGGCGGGGT	783	
Db	6522	ACTATACTACTAAACAGACCGCAACCTCAACACCACTTCTTCGACCCGCGGAGAGGA	6581	
QY	784	GACCGGATCTGTACAAACATCCTGTGGTCTTTGGGCACCCGGAAGTGTACATCATC	843	
Db	6582	GACCCCATCTATACCAACACTATTCTGATTTTCGGTCACCCCTGAATTTATATCTTT	6641	
QY	844	ATTCTGCCGCGTTTGGCATCATAGCCATGTGCTGTCGACCTTCTC--GAAAAAGCCG	900	
Db	6642	ATCCTACAGGCTTCGAAATAATCTCCATATTGTAACTTACTACTCGGAAAAAAGAA	6701	
QY	901	GTCTTCGGTTACCTGCCGATGGTCTATGCAATGGTGGCAATCGGTGTTCTGGGCTTGT	960	
Db	6702	CCATTTGGATACATAGGTAAGTGTGAGCTATGATATCAATGGCTTCTTAGGGTTTATC	6761	
QY	961	GTCTGGGCGCACCATGTACACGGTTGGTATGTTCGTGACCCAGCAATCCTACTTTCATG	1020	
Db	6762	GTGTGAGCACACACATATTTACAGTAGGAATAGACGTAGACACAGAGCATATTTCAAC	6821	
QY	1021	CTGGCCACCATGGTGTATCGGGTGCAGCCGGAATTAAGATCTTCTGTGGATTCGCCACG	1080	
Db	6822	TCCGCTACCAATAATCATCGTATCCCCACCGCGTCAAAATATTAGTACCTACCTCGCAC	6881	
QY	1081	ATGTGGGGCGGCTCGGTGTAGTTCAAAATCCCGATGCTCTGGGCGTTTGGCTTATGTTTC	1140	
Db	6882	CTCCAGGAAGCAATATGAATGATCTGTGCACTGCTGTAGCCCTAGGATTCATCTTT	6941	
QY	1141	CTGTTTCACTGGTGGTGTGACCGGTATCGTGTGGCCGCAAGCGGGTCTGGACCGTGA	1200	
Db	6942	CTTTTCACTGGTGGGCTGACTGGCAATTTGTTATTAGCAAACTCATCACTAGACATGTA	7001	
QY	1201	TATCACGACACCTATTACGTGGTGGCGCACTTCCATTTATGTGATGTCGCTGGGTGGGATC	1260	
Db	7002	CTACACGACGACTACTACGTTGTAGCCCACTTCCACTATGTCTCTATCAATAGAGCTGTA	7061	
QY	1261	TTTTCGATCTTCGCGGTATCTACTTTTATCATGCCGAAGTCTCTGGGCGCGGTTTCCCG	1320	

Db	7062	TTTGGCATCATAGGAGCTTCATTCACATGATTTCGCCCTATTCTCAGGCTACACCCTAGAC	7121
Qy	1321	GAATGGGCTGCAAAGCTGCACCTTCCTGGACCTTCCTCATCGGTGCGAAAGTCAAGTTCTTC	1380
Db	7122	CAAACTACGCCAAAATCCATTTCACATATCATATTATCATCGGCGTAAATCTAACTTCTTC	7181
Qy	1381	CCGACGACTTCCTGGGAGCGTCAGGGTATGCCGCGCGCTTACATCGACTATCCGAAGCC	1440
Db	7182	CCACAACACTTTCCTCGGCGCTATCCCGAATGCCCGACGTTACTCGGACTACCCCGATGCA	7241
Qy	1441	TTCCGCGCTTGGAAACAAAAGTCCTGCTTATGGTGGGTCTTCGGCCCTTCGGCTCGTCTCTG	1500
Db	7242	TACACCACATGAACATCCTATCACTCTGPAAGCTCATTTCATTCTCTAACAGCAGTAATA	7301
Qy	1501	TTCTTCATCGTGAATCTTT	1518
Db	7302	TTAATAATTTTCATGATT	7319

RESULT 15

```

US-10-308-264-529
; Sequence 529, Application US/10308264
; Publication No. US20040029133A1
; GENERAL INFORMATION:
; APPLICANT: Herinstdadt, Corinna
; TITLE OF INVENTION: MITOCHONDRIAL DNA POLYMORPHISM
; FILE REFERENCE: 660088.461
; CURRENT APPLICATION NUMBER: US/10/308,264
; CURRENT FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 697
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 529
; LENGTH: 16569
; <211> conflict
; TYPE: DNA
; ORGANISM: Homo sapiens mitochondrial
; FEATURE:
; LOCATION: 3107
; OTHER INFORMATION: "n" refers to a nucleotide at position 3107 of the
; OTHER INFORMATION: Cambridge Reference Sequence (CRS, Anderson et al.
; OTHER INFORMATION: 1981 Nature 290:457-465), later determined to be
; OTHER INFORMATION: absent from the human mtDNA sequence according to
; OTHER INFORMATION: the revised CRS (rCRS, Andrews et al. 1999 Nature
; OTHER INFORMATION: Genetics 23:147); "n" therefore represents a
; OTHER INFORMATION: single nucleotide deletion at a position corresponding to
; OTHER INFORMATION: nucleotide 3107 of SEQ ID NO:1 as provided herein.
US-10-308-264-529

```

	Query Match	23.8%;	Score 398.8;	DB 16;	Length 16569;
	Best Local Similarity	58.2%;	Pred. No. 5e-110;		
	Matches 744;	Conservative 0;	Mismatches 522;	Indels 12;	Gaps 2;
Qy	244	AACGGACACCTGGGAACGTCATGGTTTACCTACCACTGGTATTCTCATGATGTTCTTTGTG	303		
Db	6051	AACGACACCATCTACAAAGTTATCGTCACAGCCCATGCAATTGTGAATAATCTTCTTCATA	6110		
Qy	304	GGTATCCCGCATGTTGTCGTGGTTTGGTAACTATCTGATCCCGCTGCAAAATCGCGGT	363		
Db	6111	GTAATACCAATCATATCGAGAGCTTTGSCAACTGACTAGTTCGCCCTAATAATCGTGCC	6170		
Qy	364	CCGGATATGCGCTTCCCGGTATGAACAACCTGTGCTGTGCTGCTTCAATTCGCGGTACC	423		
Db	6171	CCGGATATGCGGTTTCCCGCATAAACAACATAAGCTTCTGACTCTTACCTCCCTCTCTC	6230		
Qy	424	GCATGGGCGTGGCTTCGCTGTTCGACCGGGCGGTGACGGTCAGCTGGGTTTCGGGCGTT	483		
Db	6231	CTACTCTGCTGCATCTCTCTATAGTGGAGCGCGAGCAGGAACAGGTTGAACAGTC---	6287		
Qy	484	GGTTGGGTTCTGTATACCGCGCGTGTCCACCCCGAAGCTGGGCTATTTCATGGACCTCCGG	543		
Db	6288	-----TACCTCCCTTAGAGGGAACACTACTCCACCCCTGGAGCCTCGTAGACCTACC	6341		
Qy	544	ATTTTCGCGGTTCACTTGTGGTGCCTCCTCGATCATGGGCGCGATCAACATGATCAG	603		

Db	6342	ATCTTCTCTTTACACCTAGCAGGTGTCCTCTATCTTAGGGCCATCAATTTCACTACA	6401
Qy	604	ACCTTCTTGAAATGCGCGCCCGGCATGACGTGCACAAAGTGCGTTGTTCTCGTGG	663
Db	6402	ACAATTATCAATATAAAACCCCTGCCATAAACCAATACCAAACGCCCTCTCGTCTGA	6461
Qy	664	TCGATCTTTATCAGCGTTGGCTGATCCTCGTGGCGCTGCGGTTCTGGGTGTTGGAATC	723
Db	6462	TCCGTCTAATACAGCAGAGTCTACTTCTCCTATCTCTCCAGTCTAGCTGTGGCATC	6521
Qy	724	ACCATGTGTGACCGACCGTAACTTCGGCACACCTTCTTCAATCTGTGCGGCGGT	783
Db	6522	ACTATACTATAAGACCGCACTCAACACACCTTCTTCGACCCGCGGAGAGGA	6581
Qy	784	GACCCGATTCTGTACCAACATCCTGTGTTCTTTGGGACCCGGGAAGTGTACATATC	843
Db	6582	GACCCCATTTATACCAACACCTATTCTGATTTTTTCGTCACCTGAAGTTTATATCTT	6641
Qy	844	ATTCTGCCCGCTTTGGCATCATCAGCCATGTCGTGTCGACCTTCTC--GAAAAAGCCG	900
Db	6642	ATCTACAGGCTTCGGAATAATCTCCATTTGTAATTACTTACTCTCGGAAAAAAGAA	6701
Qy	901	GTCTTTCGGTTACCTGCCGATGGTCTATGCAATGGTGGCAATCGGTCTTCGGCTTTGTC	960
Db	6702	CCAATTCGATACATAGGTATGGTCTGAGCTATGATATCAATTGGCTCTCTAGGTTTATC	6761
Qy	961	GTCTGGGGGACACATGTACACGTTGGTATGTCGCTGACCCAGCAATCCTACTTCAATG	1020
Db	6762	GTGTGAGCACACCATATTTACAGTAGAATAGACGTAGACACACGAGCATATTTCAAC	6821
Qy	1021	CTGCCCCACCTGGTATCGGGTCCGCACCGGATTAAGATCTTCTCGTGGATCGCCACG	1080
Db	6822	TCCGCTACCATATCATCGCTATCCCAACCGGCTGAAGTATTTAGCTGACTCGCCACA	6881
Qy	1081	ATGTGGGGGGCTCGTTGAGTTCAAAATCGCCGATGCTCTGGGCTTTGGCTTTATGTTT	1140
Db	6882	CTCCACGGAAGCAATATGAAATGATCTGCTCAGTGTCTGAGCCCTAGGATTCATCTTT	6941
Qy	1141	CTGTTTCACGGTGGTGTGACCGGTATCGTGTGCGCCCAAGCGGTCTGGACCGTGCA	1200
Db	6942	CTTTTCACGTTAGTGGCTGACTGGGATTTGTATTAGCAAACTCATCACTAGACATGTA	7001
Qy	1201	TATCAGCACACCTATTACGTGGTGGCCACTTCCATTTATGTATGTGCTGGGTGGCATC	1260
Db	7002	CTACACACACGTACTACGTTGTAGCCACTTCCACTATGTCTCTATCAATAGGAGTGTA	7061
Qy	1261	TTTCGGATCTCGCGGTATCTACTTTTACATGCGGAAGTTCTCGGCGCGCTTTCCCG	1320
Db	7062	TTTGCCATCATAGAGGCTTCATTCACTGATTTCCCTATTCTCAGGCTACACCTAGAC	7121
Qy	1321	GAATGGGCTGAAAAGCTGCATTTCTGGACCTTTCTTCATCGTGGCGAAGTCAGCTTCTC	1380
Db	7122	CAACCTAGGCCAAATCCATTTTCATATCATATTCATCGGCGTAAATCTAACTTCTTC	7181
Qy	1381	CCGAGACATTTCTGGAGCTCAGGGTATGCGCGCGCTTACATCGACTATCCGAAGCC	1440
Db	7182	CCACAACACTTTCTCGCGCTATCCGGAATGCCGACGTTACTCGGACTACCCCGATCA	7241
Qy	1441	TTCCGCTGTGGAAACAAAGTCTCGTCTATGTTGCTCTCTGGCTTCGCTCGTCTG	1500
Db	7242	TACACCATGAACATCCTATCATCTGTAGGCTCATTTCTCTAAACAGCAATA	7301
Qy	1501	TTCTTCATCGTGATCTTT	1518
Db	7302	TTAATAATTTTCATGATT	7319

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 22, 2004, 14:17:20 ; Search time 7629 Seconds
(without alignments)
7995.820 Million cell updates/sec

Title: US-09-712-768A-1
Perfect score: 1674
Sequence: 1 atggcagacgcgcattca.....acaagcatccctgcactaa 1674

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_gss1:*
8: gb_gss2:*
9: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	411.4	24.6	1515	CR696017 Tetraodon
2	409.8	24.5	1528	CR701206 Tetraodon
3	408.2	24.4	1614	CR694572 Tetraodon
4	408.2	24.4	1619	CR688040 Tetraodon
5	408.2	24.4	1638	CR704060 Tetraodon
6	406.6	24.3	1642	CR691904 Tetraodon
7	403.4	24.1	1486	CR680278 Tetraodon
8	403.2	24.1	1590	CR695115 Tetraodon
9	397.8	23.8	1426	CR689102 Tetraodon
10	397.2	23.7	3167	BC024102 Homo sapi
11	396.2	23.7	1486	CR597319 Tetraodon
12	396.2	23.7	1499	CR590798 Tetraodon
13	396.2	23.7	1500	CR598101 Tetraodon
14	396.2	23.7	1504	CR615203 Tetraodon
15	396.2	23.7	1510	CR623733 Tetraodon
16	396.2	23.7	1543	CR626364 Tetraodon
17	396.2	23.7	1546	CR601538 Tetraodon
18	396.2	23.7	1547	CR616898 Tetraodon
19	396.2	23.7	1552	CR623180 Tetraodon
20	396.2	23.7	1556	CR602727 Tetraodon
21	396.2	23.7	1559	CR590922 Tetraodon
22	396.2	23.7	1559	CR595406 Tetraodon
23	396.2	23.7	1562	CR618826 Tetraodon
24	396.2	23.7	1563	CR626271 Tetraodon

25	396.2	23.7	1567	3	CR595999	full-leng
26	396.2	23.7	1567	3	CR605797	full-leng
27	396.2	23.7	1570	3	CR598837	full-leng
28	396.2	23.7	1570	3	CR599661	full-leng
29	396.2	23.7	1574	3	CR626148	full-leng
30	396.2	23.7	1576	3	CR604333	full-leng
31	396.2	23.7	1576	3	CR609160	full-leng
32	396.2	23.7	1577	3	CR624333	full-leng
33	396.2	23.7	1585	3	CR623432	full-leng
34	396.2	23.7	1586	3	CR612356	full-leng
35	396.2	23.7	1597	3	CR624374	full-leng
36	396.2	23.6	1578	3	CR598056	full-leng
37	395.6	23.6	1578	3	CR732481	Tetraodon
38	394.6	23.6	1425	3	CR682179	Tetraodon
39	392.8	23.5	1421	3	CR685111	Tetraodon
40	392.6	23.5	1591	3	CR697573	Tetraodon
41	390.2	23.3	1505	3	CR62973	Tetraodon
42	388.8	23.2	1433	3	CR704783	Tetraodon
43	385.8	23.0	1532	3	CR698071	Tetraodon
44	384.8	23.0	1437	3	CR727707	Tetraodon
45	383.4	22.9	1443	3	CR600831	full-leng

ALIGNMENTS

RESULT 1
CR696017
LOCUS CR696017 1515 bp mRNA linear HTC 12-AUG-2004
DEFINITION Tetraodon nigroviridis full-length cDNA.
ACCESSION CR696017
VERSION CR696017.1 GI:5193924
KEYWORDS HTC; cDNA; full-length; Tetraodon nigroviridis.
SOURCE Tetraodon
ORGANISM Tetraodon
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodontidae.
REFERENCE 1 (bases 1 to 1515)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (10-AUG-2004) Genoscope - Centre National de Sequencage - 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
COMMENT (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
The sequences are based on single pass reads.
More information available at
http://www.genoscope.cns.fr/tetraodon.

FEATURES
Location/Qualifiers
source 1..1515
/organism="Tetraodon"
/mol_type="mRNA"
/db_xref="taxon:47144"
/tissue_type="Eyes"
ORIGIN
Query Match 24.6%; Score 411.4; DB 3; Length 1515;
Best Local Similarity 58.9%; Pred.No. 2.4e-99;
Matches 750; Conservative 0; Mismatches 511; Indels 12; Gaps 2;
QY 248 GACACCTGTGGAACGTCATGTTACCTACCATGGTATCTGATGATGTTCTTTGGGTA 307
Db 61 GCCAGATTATAAGTAATGTTACGCCACCGCATTTGTAATAATTTCTTTATAGTAA 120
QY 308 TCCCGGATTTGTCGGTGGTTTGGTAACATCTGATGCCGCTGCAATCGCGCTCGG 367
Db 121 TGCCAATCATGATGGCGGCTTTGGAACATGACTAGTGGCCCTCATGTTGGACACCTG 180
QY 368 ATATGGCTTCCCGGCTATGAACACCTGTCGTTCTGGCTGTTTCATTCGCGGTACCGCGA 427
Db 181 ACATGGCATTTCCCGGATATAAATACATAGCTTTTGGTTACTACCTCCCTTCCTCTAC 240
QY 428 TGGCGTGGCTTCTGCTGTTCGACCGGGGTGACCGGTTCAGCTGGTTCGGCGGTTCGT 487

```
Db      241  TCCTCTAGCAT-----CCTCCGGCTAGAAGCTGGGGCTGTACAGGCTGAACCG 291
QY      488  GGGTTCTGTACCGCGCTGTCGACCGCGAAGCTGCTATTTCGATGGACCTCGCGATT 547
Db      292  TCTACCCCACTAGCAGGCAACTGCGCCACGCGGAGCTTCGGTTGACCTAACCATCT 351
QY      548  TCGCGGTTCACTTGTTCGGGTGCTCTCGATATGCGCGCATCAACATGATCAGCACT 607
Db      352  TCTCCCTCCACCTGGCGCGTGTCTATCGATTATAGGGCCATTAACTTTATTACCACA 411
QY      608  TCTTGACATGCGCGCGCGGATGATGCTGACAGCTGCACAAAGTGGCTTCTCGTGTGA 667
Db      412  TCATTAACTAAACCCCGAGCATCTCCCAATATCAGGCCCTCTCTTCGTATAGCCG 471
QY      668  TCTTTATCAGGCTTGGCTGATCTGCTGGCGTCTGCGGTTCTGGGTGTCGAATCACA 727
Db      472  TCTTAATTACTGCGCTCTCTCTGCTGCTGCTGCGAGTCTCGCAGCGGATCACAA 531
QY      728  TGTGCTGACGACGTAATCTGCGACGACCTTCTTCAATCTGCTGCGCGCGGTGACC 787
Db      532  TGTCTCTGACAGACCGAACTTAACACCACTTCTTTCGACCTGCGCGGAGGGACC 591
QY      788  CGATTCTGTACCAACACATCTGTGGTTCCTTGGGACCGGAGTGTATCATCATCTC 847
Db      592  CAATCTTTACCAACACTTATCTGATCTTTGGCCACCGCAGAGTGTATCTCATCC 851
QY      848  TGCCCGCTTTGGCATCATCAGCATGTGTGTGACCT--TCTGAAAGCCGGTCT 904
Db      652  TACCAGGCTTTGAAATAATCTCCCATATCTGCTGCTATGTCGCGCAAAAGAACCAT 711
QY      905  TCGGTTACCTGCGATGCTATGCAATGCGCAATCGGTGTTCTGGGCTTTGCTGCT 964
Db      712  TCGGCTACATGGGATGGTTTGAGCCATGATGGCTATGSCCTTCTGGGCTTCATGCT 771
QY      965  GGGCGCACCATGTATACCGCTTGGTATGTCGTGACCCAGCAATCTATTCATGCTGG 1024
Db      772  GAGCCCAACCATATGTTTACGGTAGGAATGACGCTTGACACCGCGCTACTTCACATCG 831
QY      1025  CCACCTGCTGTCGCGTCCGACCGCATTAAGATCTCTCGTGATCGCCACGATGT 1084
Db      832  CCACATAATTAATGCCATCCCCACCGAGTAAAGTATTAAGTACTAGCAGCACCCTTC 891
QY      1085  GGGCGGCTCGGTGAGTTCAAATCGCGGATGCTCTGGGCTTTGGCTTTATCTCTCTG 1144
Db      892  ACGCGGCTCCATTAAATGAGAACTCCAAATGCTCTGAGCCCTAGGCTTCATCTCTCT 951
QY      1145  TCACGCTGGTGGTGTACCGGTATGCTGTGGCCCAAGCGGTCTGGACCGTGATATC 1204
Db      952  TTACCGTAGCGGACTAACAGGCAATGCTTAGCCAACTCTCCCTAGACGTAGTCTCTG 1011
QY      1205  ACGACACTTATAGTGGTGGGCACTTCATTAATGATGATGCTCGCTGGGTGCGATCTTG 1264
Db      1012  ATGACACTTACTAGTAGTTGCCACTTCCACTACGTCCTATCAATAGAGCGGATTCG 1071
QY      1265  CGATCTTGGCGGTAATCTATTTTACATGCGGAAGTTCGGCGCGGCTTTCCCGAAT 1324
Db      1072  CTATCATGGGTGCTTTTGTACATGATTCACCTTTTCTCAGGGTACACCTCCACAGCA 1131
QY      1325  GGGCTGCAAGCTGCACTCTCGACCTTCTTCACTCGGTGCGACGTCAGTCTTCCCGC 1384
Db      1132  CTTGAAACAAATCCATTTTGGAGTTATGTTCTAGGCGTTAACTTAACCTTCTCCCGC 1191
QY      1385  AGCACTTCTTGGGAGCTGAGGATGCGCGCGGTTACATCGACTATCCGAGCGCTTCG 1444
Db      1192  AACACTTCTTGGAGTGTGCGGATGCGCGTCTGATCTCAGACTACCCAGACGCGTACA 1251
QY      1445  CGGTGTGGAACAAAGTCTGCTCTATGGTGTGGTCTGCGCTTCTGCTTCTCTCTCT 1504
Db      1252  CGCTATGAAACACAGTTTCTTCGATGGCTCTCTAAATCTCACTATCGCTGTCTATAT 1311
QY      1505  TCATCGTGTACTTT 1517
```

```
Db      1312  TCCTATTATCCT 1324
RESULT 2
CR701206
LOCUS      CR701206      1528 bp      mRNA      linear      HTC 12-AUG-2004
DEFINITION Tetraodon nigroviridis full-length cDNA.
ACCESSION CR701206
VERSION   CR701206.1 Gi:51199115
KEYWORDS  HTC; cDNA; full-length; Tetraodon nigroviridis.
SOURCE    Tetraodon
           Tetraodon
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
           Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
           Tetraodontidae; Tetraodontidae.
REFERENCE 1 (bases 1 to 1528)
AUTHORS   Direct Submission
TITLE     Submitted (10-AUG-2004) Genoscope - Centre National de Sequencage -
JOURNAL   : 2 rue Gaston Cremlieux, CP 5706 - 91057 EVRY cedex - FRANCE
           (E-mail : segre@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT   The sequences are based on single pass reads.
           More information available at
           http://www.genoscope.cns.fr/tetraodon.
FEATURES  Location/Qualifiers
           source
           1..1528
           /organism="Tetraodon"
           /mol_type="mRNA"
           /db_xref="taxon:47144"
           /tissue_type="Eyes"
ORIGIN
Query Match      24.5%; Score 409.8; DB 3; Length 1528;
Best Local Similarity 58.7%; Pred. No. 6.4e-99;
Matches 752; Conservative 0; Mismatches 517; Indels 12; Gaps 2;
QY      240  GGCAGACCGACACCTGTGGAACTCATGGTTACCTACATGGTATTCTGATGATGTTCTT 299
Db      25   GGGTCATGACAGATTTATATGTAATCGTTACAGCCCATGCAATTGTAATATTTCTT 84
QY      300  TGTGGGTATCCCGCATGTTTCGGTGGTTTGGTAACTATCTGATGCGGCTGCAATCGG 359
Db      85   TATAGTAATGCAATCATGATTGGCGCTTTGGAACCTGACTAGTGCCTCATGATTGG 144
QY      360  CGCTCCGGATATGCGCTTCCCGCTATGACAACTGCTGCTTCTGGCTGTTTCATTCGCG 419
Db      145  AGCACCTGACATGCAATTCGCCCAATTAATACATAAGCTTTGGTACTACCTCCTTC 204
QY      420  TACCGCATGGCGTGGCTTCGCTGTTGCGACCGGCGGTGACGGTCAGTGGGTTCCGG 479
Db      205  CTTCTACTCTCTCTAGCAT-----CCTCCGGGTAGAACTGGGCTGGTACAGG 255
QY      480  CGTTGGTTGGGTTCTGTACCGCGCTGTGACCCGCGAAGCTGGCTATTTCGATGGACCT 539
Db      256  CTGAACCGTCTACCCCGCATAGCAGGCAACCTGGCCCAACGCGGAGCTTCGGTTGACCT 315
QY      540  CGCATTTTCGCGGTTCACTTTGTCGGGTGCTCTCGATCATGCGCGCGCATCAACATGAT 599
Db      316  AACCATCTTCTCCCTCCACCTGCGGCTGTCATCGATTTTAGGGCCATTAACCTTAT 375
QY      600  CAGACCTTCTTGAACATGCGCGCGCGCATGACCTGACCTGACAAAGTGGCTGTCTC 659
Db      376  CACCACCATATTAAATAAAACCCCGAGCATCTCCCAATATCAGACCCCTCTCTTCGT 435
QY      660  GTGTCGATCTTATACGGGTTGGCTGATCCTCTGCGCTGCGCTGCGGCTTCTGGCTGTGC 719
Db      436  ATGAGCGCTCTAATTAAGTGGCTCTTCTCTACTGTCCTTGGCAGTCTCTCGAGCGCG 495
QY      720  AATCACCATGCTGTGACCGGCTTAATCTTGGGACGACCTTCTTCAATCTGCTGCGGG 779
Db      496  GATCACAATGCTTCTTAACAGACCGGAAACCTTAACACCACTTCTTTCACCTCTGAGCGG 555
```



```
QY 780 CGGTGACCGGATCTGTACCAACACATCCTGTGGTCTTTTGGGACCGGAGGTGATCAT 839
Db 556 AGGAGACCCAAATCTTTACCAACATTAATCTGTATCTTTGGCCACCCAGAGTGTACAT 615
QY 840 CATCATCTGCGCGGCTTTGGCATCATCAGCCATCTCGTGTGACCTTCTCG---AAAA 896
Db 616 TCTCATCTTACAGGCTTTGGAATTAATCTCCCATATTTGTGCTACTATGCGGCAAAA 675
QY 897 GCGGTCTTTCGGTTACCTGCGGATGGTCTATGCAATGGTGGCAATCGTGTCTGGGCTT 956
Db 676 AGAACCATTCGGCTACATGGGCATGGTTGGGCCATGATGGCTATTGGCTTCTAGGCTT 735
QY 957 TGTGCTGCGGCGGACCAACATGTACACGTTGGTATGTCGTACCGACCATCTACTT 1016
Db 736 CATGCTGTAGCGCCACCATATGTTACGGTAGGAATGGAAGTGTGACACCGGCGCTACTT 795
QY 1017 CATGCTGGCCACATGTTGATCGCGGTGCCGACCGGCATTAAGATCTTCTGTCGATCGC 1076
Db 796 CACATCCGCCCAATTAATTTATGGCCATCCCAACCGAGTAAAGTATTAGCTACTAGC 855
QY 1077 CACGATGCGGCGGCTGGGTTAGTTCAAATCGCGGATGCTCTGGGCTTTGGCTTTAT 1136
Db 856 CACCCTTACCGCGGCTCCATTAATAATGAGAACTCCAATGCTCTGAGCCCTAGGCTTCAT 915
QY 1137 GTTCTGTTTACCGTGGGTGGTGTGACCGGTATCGTGTGGCCCAAGCGGCTCTGACCG 1196
Db 916 CTTCTCTTTACGTAGCGGACTAACAGCATTTGCTTAGCCAACTCTCTCCCTAGACGT 975
QY 1197 TGCATATACGACACTATTACGTGGTGGCGCACTTCCATCTTTCATCGGTGCAACGTGCG 1256
Db 976 AGTCTGTGATGACACCTACTAGTATGTCCTGACCTTCCACTACGTCCTATCAATAGGAGC 1035
QY 1257 GATCTTTGGATCTTGGCGGTTATCTATTTTACATGCGGAAGTCTTGGGCGGCTTT 1316
Db 1036 CGTATTGTCATCATAGGTCCTTTGTACATGATTCCTCACTTTCTCAGGATACACCT 1095
QY 1317 CCGGAATGGGCTGCAAGCTGACCTTCTGACCTTCTTTCATCGGTGCAACGTGACGTT 1376
Db 1096 CCACAGCACTGAAACAAAATCCATTTGGAGTTATGTTGCTAGGCGTTAACTTAACCTT 1155
QY 1377 CTTCCGGAGCACTTCTGGGAGGTGAGGTATGCGGCGCTTACATCGACTATCCCGA 1436
Db 1156 CTTCCCGCAACACTTCTCGGACTCGCGGAATGCGCGCTGATACATCAGATTACCCAGA 1215
QY 1437 AGCTTTCGGCTGTGGAACAAAGTCTGCTCTATGTTGCTGCTTCTGCGCTTCTGCTCGT 1496
Db 1216 CGGTACAGCTATGAAACACAGTTTCTCTCATTTGCTCTCTAACTCATCTATCGCTGT 1275
QY 1497 CTTGTTCTTCATCGTGAATCTT 1517
Db 1276 CATTATATTCCTATTATCCT 1296
```

RESULT 3

```
CR694572
LOCUS Tetraodon nigroviridis 1614 bp mRNA linear HTC 12-AUG-2004
DEFINITION Tetraodon nigroviridis full-length cDNA.
ACCESSION CR694572
VERSION CR694572.1 GI:51192479
KEYWORDS HTC; cDNA; full-length; Tetraodon nigroviridis.
SOURCE Tetraodon
```

ORGANISM

Tetraodon
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae.

REFERENCE

```
1. (bases 1 to 1614)
Genoscope.
Direct Submission
Submitted (10-AUG-2004) Genoscope - Centre National de Sequencage -
: 2 rue Gaston Crémieux, CP 5706 - 91057 EVRY cedex - FRANCE
(E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
```

COMMENT

The sequences are based on single pass reads.

More information available at
<http://www.genoscope.cns.fr/tetraodon>.

FEATURES

source
1..1614
/organism="Tetraodon"
/mol_type="mRNA"
/db_xref="taxon:47144"
/tissue_type="Eyes"

ORIGIN

```
Query Match 24.4%; Score 408.2; DB 3; Length 1614;
Best Local Similarity 58.6%; Pred. No. 1.8e-98;
Matches 751; Conservative 0; Mismatches 518; Indels 12; Gaps 2;

QY 240 GCGGAACGACACCTGTGGAAAGTCATGTTACTACCATGTTATCTGATGCGGTGCAAAATCGG 359
Db 135 GGGTGATGACAGATTTATAATGTAATCGTTACAGCCACGCATTTGTAATAATTTTCTT 194
QY 300 TGTGGGTATCCCGCATTTGCGTGGTGGTAACTATCTGATGCGGTGCAAAATCGG 359
Db 195 TATAGTAATGCAATCAATGATTTGGCGCTTTGAAACTGACTAGTCCCTCATGATGG 254
QY 360 CGTTCGGATATGCGCTTCCCGGTATGAACAACTGTCTGTTCTGGGTGTTCAATGCGGG 419
Db 255 AGCACCTGACATGGCATTTCCCGGATAAATAACATAAGCTTTTGGTTACTACCTCCTTC 314
QY 420 TACCGCATGGCGTGGCTTGGCTGTTTCGCACCGGCGGTGACGTCAGCTGGTTCGGG 479
Db 315 CTTCTACTCTCTTAGCAT-----CCTCGGCGTAGAAGCTGGGCTGGTACAGG 365
QY 480 CGTTGGTTGGGTTCTGTACCGCGCGTGTGCAACCGGAAAGCTGGCTATTTCGATGGACCT 539
Db 366 CTGAACCGTCTACCCGCCACTAGCAGGCAACCTGGGCCACCGCGAGCTTCGGTTGACCT 425
QY 540 CGCATTTTCGGGTTCACTTGTGGGTGCTCTCGATCATGGCGCGATCAACATGAT 599
Db 426 AACCATCTTCTCCCTCCACCTGGCGGTGCTCATCGATTTTAGGGGCCATTAACCTTAT 485
QY 600 CACGACCTTCTGAACATGGCGGCGCCCGGCATGACGCTGCACAAAGTCGCGTTGTTCTC 659
Db 486 TACACACATCATTAACATAAACCAGCACTCTCCCAATATCAGACCCCTCTCTTCGT 545
QY 660 GTGTCGATCTTTATCAGGCTTGGCTGATCTCTGTCGCGCTGCGGTTCTGGCTGTGC 719
Db 546 ATGAGCGCTCTTAATTAATCTGCGCTCTCTCTGCTCTGCTGCTGCTGCTGCTGCTG 605
QY 720 AATCACCATGCTGTGACCGACCGTAACCTTCGGCAGCACCTTCTTCAATCTCTGTGCGG 779
Db 606 GATCACATGCTTCTGACAGACCGAAACCTTAACACACACTTCTTCGACCTTCGAGCGG 665
QY 780 CGGTGACCCGATCTGTACCAACACATCTGTGTGTTCTTTGGGACCCGGAAGTGTACAT 839
Db 666 AGGGAGCCCAATCTTTTACCAACACTTATTCTGATTCTTTGGCCACCCAGAGTGTACAT 725
QY 840 CATCATTTGCGCGCTTTGGCATCATCAGCCATGTGCTGTGACCTTCTC---GAAAAA 896
Db 726 TCTCATCTACAGGCTTTGGAATAATCTCCCATATTGTTGCTCTATTATGCGCGCAAAA 785
QY 897 GCGGCTCTTCGGTTTACCTGCGATGGTCTATGCAATGGTGGCAATCGGTGTTCTGGGCTT 956
Db 786 AGAACCATTCGCTACATGGGCATGGTTTGAGCCATCATGGCTATTGGCTTCTGGGCTT 845
QY 957 TGTGCTGTGGCGGACACATGATACCGTTGGTATGCTGCTGACCCAGCAATCTTACTT 1016
Db 846 CATGCTGTAGCCCAACCATATGTTTACGGTAGGAATGAGCTTGACACCCCGCGCTACTT 905
QY 1017 CATGCTGGCCACCATGTTGATGCGGTTGCGGCGCATTAAGATCTTCTGTCGATCGC 1076
Db 906 CACATCCGCCACATAATATTGTCATCCCAACCGAGTAAAGTATTAGCTACATAGC 965
QY 1077 CACGATGTGGCGGCTCGGTGAGTTGAGTTCAAAATCGCGGATGCTTGGGCTTTTGGCTTAT 1136
Db 966 CACCTTTCACGCGGCTCCATTAATAATGAGAACTCCAATGCTGCTGAGCCCTTAGCTTCAT 1025
```


Db 1332 CGGTCACAGCTATGAACACAGATTTCCTCCATTGGCTCTCTAAATCTCACTTATCGCTGT 1339
 Qy 1497 CTTGTTCTTCATCGTGATCTT 1517
 Db 1392 CATTATATTCCTATTTATCCT 1412

RESULT 5
 CR704060 1638 bp mRNA linear HTC 12-AUG-2004
 LOCUS Tetraodon nigroviridis full-length cDNA.
 DEFINITION CR704060
 ACCESSION CR704060.1 GI:51201969
 VERSION HTC; cDNA; full-length; Tetraodon nigroviridis.
 KEYWORDS Tetraodon
 SOURCE Tetraodon
 ORGANISM Tetraodon
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodontidae.
 1 (bases 1 to 1638)
 Genoscope.
 Direct Submission
 TITLE Submitted (10-AUG-2004) Genoscope - Centre National de Sequencage
 JOURNAL : 2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
 COMMENT (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 The sequences are based on single pass reads.
 More information available at
 http://www.genoscope.cns.fr/tetraodon.
 FEATURES:
 source location/Qualifiers
 1..1638
 /organism="Tetraodon"
 /mol_type="mRNA"
 /db_xref="taxon:47144"
 /tissue_type="Eyes"

ORIGIN
 Query Match 24.4%; Score 408.2; DB 3; Length 1638;
 Best Local Similarity 58.6%; Pred. No. 1.8e-98;
 Matches 751; Conservative 0; Mismatches 518; Indels 12; Gaps 2

Qy 240 GSCGAAGGACACCTGTGGAAGCTCATGTGTACCTACCATGCTATTCTGATGATTTCTT 299
 Db 141 GGGTGATGACCAAGATTATAATGTAATCGTTACGCCACGCAATTGTAATAATTTCTT 200
 Qy 300 TGTGGGTATCCCGGATTTCTTCGGTGGTTTGGTAACTATCTGATGCCGCTCCAATCGG 359
 Db 201 TATAGTAATGCCAATCATGATTGGCGCTTTGGAAACTGACTAGTGCCTCATGATTGG 260
 Qy 360 CGCTCCGGATATGGCTTCCCGGTATGAACACCTGTGTTCTGCTGTTCAATTGCCGG 419
 Db 261 AGCACCTGACATGGCAATCCCCCGAATAAATAACATAAGCTTTGGTTACTACCTCTTC 320
 Qy 420 TACCGCGATGGCGTGGCTTCGTGTTCGACACGGCGGTGACGGTCACTGGTTCGGG 479
 Db 321 CTTCTACTCCCTCTAGCAT-----CCTCCGCGGTAGAAGCTGGCGCTGTACAGG 371
 Qy 480 CGTTGGTTGGGTTCTGTATACCCGCGCTGTGCAGCCCGGAAGCTGGCTATTTCGATGACCT 539
 Db 372 CTGAACCGCTCTACCCCCCACTAGCAGCAACCTTGGCCCCACGCGGAGCTTCGGTTGACCT 431
 Qy 540 CGGATTTTCGCGGTTTCACCTFTGCGGTGCTCTCTCGATCATGGGCGCGCATCAACATGAT 599
 Db 432 AACCATCTTCTCCTCCACTGCACCTGCGCGGTGTCCTCATCGATTTAGGGGCCATTAACTTAT 491
 Qy 600 CACGACCTCTTGAACATCGCGCCCGCCGCGCATGACGCTGCACAAGTCCGTTGTCTTC 659
 Db 492 TACCACCATCATTAACATAAAACCCCGAGCAATCTCCCAATATCAGACCCCTCTCTTCGT 551
 Qy 660 GTGTCGATCTTTATCACGGCTTGGCTGATCTCTGCTGGCGCTGCCGTTCTCGCTGTGTC 719
 Db 552 ATGAGCGGCTTTAAATTAATCGCGTCTTCTTCGCGCTCTTCGCGAGTCTCCGAGCCGG 611

Qy	720	AATC	CGATGCTGCTGCACGACGGTAACCTCGGCAGACCTTCTTTCAACTCCTGCTGGCGG	779
Db	612	GATCA	CAATGCTTCTGACAGACCGAAACCTAAACCACTTCTTTCGACCTTGACGGCGG	671
Qy	780	CGGTG	ACCGGATTCCTGTACCAACACATCCTGTGGTTCTTTGGGCAACCGGAAGTGATCAT	839
Db	672	AGGG	ACCCAAATCTTTTACCAACACTTATCTGATCTTTGGCCACCCAGAAAGTGATCAT	731
Qy	840	CATCAT	TCTGCCCGGCTTTGGCATCATGACCCATGTCGTGTCGACCTTCTC--GAAAAA	896
Db	732	TCTCAT	CTCACCAGGCTTTGGATAATATCTCCCATATTTGTGGCTATTTATCGCGGCAAAAA	791
Qy	897	GCCGG	TCTCCGGTTACTGCGCATGCTCTATGCAATGGTGGCAATCGGTGTTCTGGGCTT	956
Db	792	AGAAC	CATTCGGCTACATGGGCATGGTTTGAGCCATGATGGCTATTTGGCCTTCTGGCCTT	851
Qy	957	TGTG	CTGTGGGCGACACATGATACACGGTTGTGTATGTCGCTGACCCAGGAATCTTACTT	1016
Db	852	CATCG	CTGAGCCACCATATGTTTACGGTAGGAATGGACGTTTGACACCGCGGCCCTACTT	911
Qy	1017	CATG	CTGGCACACATGTTGATCGCGGTGCCGCGCCGACCGGCATTAAGATCTCTCGTGGATCGC	1076
Db	912	CACAT	CGCCACCAATAATATTGCCATCCCAACCGGAGTAAAGATATTTAGTGTACTAGC	971
Qy	1077	CACGAT	GTGGGCGGCTCGGTTCAGTTTCAATCGCCGATGCTCTGGGCGCTTTGGCTTTAT	1136
Db	972	CACCC	TTCACGGCGCTCCATTAATCAGAAACTCCCAATGCTCTGAGCCCTAGGCTTCAT	1031
Qy	1137	GTTCT	GTGTTACCGTGGGTGTGTGACCGGTATGTCGTGCTGCCCAAGCGGGTCTGGACCG	1196
Db	1032	CTTCT	CTTTTACCGTAGCGGACCTTAACAGGCAATGTCTTAGCCAACTCTCCCTAGACGT	1091
Qy	1197	TGCAT	ATACGACACCTTATACGGTGTGGCGCATCTCCATTTGTGTATGTCGTGGGTGTC	1256
Db	1092	AGT	CCTGTATGACACTTACTACGTAGTTGCCACTTCCACTCCACTACGTCTATCAATAGGAC	1151
Qy	1257	GATCT	TTTTCGATCTTCGCGGTATCTACTTTTACATCCGAAGTTCTGGGCGCGGCTTT	1316
Db	1152	CGTAT	TCGCTATCATGGGTGCTTTGTACACTGATTCCTTCTTCAGGGTACACCCCT	1211
Qy	1317	CCGG	AATGGGCTGAAAGCTGCACTTCTGACCTTCTTCATCGTGGCAACGTCACGTT	1376
Db	1212	CCAC	AGCACTTGAACAAAAATCCATTTTGGAGTTATGTTCTGTAGCGGTTAACTAACCTT	1271
Qy	1377	CTTCC	CGCAGCACTTCTGGGACGTCAGGGTATCCGCGCGGTTTACATCGACTATCCCGA	1436
Db	1272	CTTCC	CCCCAACACTTCTCGGACTTCGCGGGAATGCCCGTCTGATCTCAGACTACCCAGA	1331
Qy	1437	AGCCT	TTCGCGTGTGGAAACAAAGTCTGTCTCTATGGTGGGTTCTTGGCCTTCGCTCGTT	1496
Db	1332	CGCG	TACACGCTATGAAACACAGATTTCTCCATTGGGCTCTCTAATCTCACTTATCGTGT	1391
Qy	1497	CTGT	TCTTTCATCGATGATCTT	1517
Db	1392	CATT	ATATCTCTATTATTCCT	1412

RESULT 6	
CR691904	
LOCUS	CR691904
DEFINITION	Tetraodon nigroviridis full-length cDNA.
ACCESSION	CR691904
VERSION	CR691904.1 GI:5119811
KEYWORDS	HTC; cDNA; full-length; Tetraodon nigroviridis.
SOURCE	Tetraodon
ORGANISM	Tetraodon
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontoideae; Tetraodontidae.
REFERENCE	1 (Bases 1 to 1642)
AUTHORS	Genoscope.
TITLE	Direct Submission

QY 665 CGATCTTTATCACGGCTTGGCTGATCTGCTGGCGCTGCCGGTCTGGCTGTGCAATCA 724
 Db 526 CCGTCTTTAAATACTGCGGCTCTTCTGCTGTCTTGGCAGTACTCGCAGCGGGATCA 585
 QY 725 CCATGCTGCTGACGACCGTAATCTCGGACGACCTCTTCAATCTGCTGGGGGGG 784
 Db 586 CAATGCTTCTGACAGACGAACTTAACACACTTCTTTCGACCTTGCAGGCGGAGGG 645
 QY 785 ACCGATCTGTACCAACACATCTGTGTTTGGGACCCGGAAGTGTACATCATCA 844
 Db 646 ACCCAATCTTTACCAACACTTATCTGATTTTGGCCACCCAGAGTGTACATCTCA 705
 QY 845 TTCTGCGCGGCTTGGCATCATGACCATCTGCTGCTGACCT---TCTGAAAAAGCCGG 901
 Db 706 TCTTACAGGCTTTGGAATTAATCTCCATATTTGCTCTTAATGCGGCAAAAAAGAAC 765
 QY 902 TCTTGGGTTACCTGCGCATGCTATGCAATGTGGCAATCGGTGTTCTGGGCTTTGTCG 961
 Db 766 CATTCGGCTACATGGCATGTTTGAGCCATGATGCTATGACCTTCTGGGTTTCATCG 825
 QY 962 TCTGGGCGCACCATGATACCGTTGGTATGCTGACCCAGCAATCTTACTTCATGC 1021
 Db 826 TCTGAGCCCAACATATGTTTACGGTAGGAATGACCTTGACACCCGCGCTACTTCCAT 885
 QY 1022 TGGCCACCATGTGTATCGCGTGGCGACCGCATTAAGATCTTCTGCTGATCGCCACGA 1081
 Db 886 CCGCCACATAATATATGCCATCTCCACCGAGTAAAGTATTTAGCTGACTAGCCACC 945
 QY 1082 TGTGGGCGGCTCGTTGAGTTCAATGCGCGATGCTCTGGGCTTTGGCTTTATGTTCC 1141
 Db 946 TTCAGCGCGCTCCATTAATAGAAACTCCMATGCTCTGAGCCCTAGGCTTCATCTCC 1005
 QY 1142 TGTTCACGTTGGTGTGACCGGTATCGTGTGGCCCAAGGGGTCTGGACCGTGCAT 1201
 Db 1006 TCCTTACCGTAGCGGACTAACAGGCAATGCTTAGCCAACTCTCCCTAGACGTAGTCC 1065
 QY 1202 ATCAGACACCTATTACGTGGTGGCGCACTTCCATATATGATGCTGGTGGGATCT 1261
 Db 1066 TGCATGAACCTTACTAGTAGTTGGCCACTTCCACTACGCTCTATCAATAGAGCGGTAT 1125
 QY 1262 TTGCGATCTTCGCGGTATCTACTTTTACATGCGGAATTTCTGGGCGCGCTTTCCCGG 1321
 Db 1126 TCCTATATGCGTGTCTTTGTACACTGATTCCTCACTTTCTCAGGCTACACCTCCACA 1185
 QY 1322 AATGGGCTGCAAGCTGCACTTCTGACCTTCTCATCGTGGCAACGTCAGTCTTCTCC 1381
 Db 1186 GCACCTGAACAAAAATCCATTTTGGAGTTATGTTCTGAGCGTTAACTTAACTTCTCC 1245
 QY 1382 CGCAGCACTTCTGGGACGTAGGATGCGCGCGGTTTACATGCACTATCCCGAAGCT 1441
 Db 1246 CCAAAACTTCTCGGACTCGCGGAATCCCGCTCGATCTCAGACTACCCAGACGGT 1305
 QY 1442 TCGCGCTGTGGAACAAAGTCTGCTCATGTGCGTTCCTGCGCTTCTGCTGTTCTGT 1501
 Db 1306 ACACGCTATGAACACAGTTTCTCATTTGGCTCTCTAATCTCACTTATGCTGTCAITA 1365
 QY 1502 TCTTCACTGTATCTT 1517
 Db 1366 TATTCCTATTATCTCT 1381

RESULT 9

CR689102
 LOCUS CR689102 1426 bp mRNA linear HTC 12-AUG-2004
 DEFINITION Tetraodon nigroviridis full-length cDNA.
 ACCESSION CR689102
 VERSION CR689102.1 GI:51187009
 KEYWORDS HTC: cDNA; full-length; Tetraodon nigroviridis.
 SOURCE Tetraodon
 ORGANISM Tetraodon
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;

Tetraodontidae; Tetraodontidae.
 1 (bases 1 to 1426)
 Genoscope.
 Direct Submission
 Submitted (10-AUG-2004) Genoscope - Centre National de Sequencage -
 : 2 rue Gaston Cremlieux, CP 5706 - 91057 EVRY cedex - FRANCE
 (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 The sequences are based on single pass reads.
 More information available at
 http://www.genoscope.cns.fr/tetraodon.
 Location/Qualifiers
 1..1426
 /organism="Tetraodon"
 /mol_type="mRNA"
 /db_xref="taxon:47144"
 /tissue_type="Eyes"

ORIGIN

Query Match 23.8%; Score 397.8; DB 3; Length 1426;
 Best Local Similarity 58.9%; Pred. No. 1.1e-95;
 Matches 728; Conservative 0; Mismatches 497; Indels 12; Gaps 2;
 QY 284 TTCTGATGATCTTCTTGTGGGTATCCCGCATTTGTTGGTGTGTTGTTGTAATCTGA 343
 Db 3 TTGTAATAATTTCTTTATAGTAATGCCAATCATGATGGCGCTTTGGAACCTGACTAG 62
 QY 344 TGGCGCTGCAAAACGGCGCTCCGGATATGGCTTCCCGCTATGAACAACTGTCTCTTCT 403
 Db 63 TGGCCCTCATGATTGGAGCACTGACATGGCAATCCCGCAATAAATAACATAAGCTTTT 122
 QY 404 GGCTGTTCAATGCGGTACCGCATGGCGTGGCTTGGCTTGGCACCGGGGGGTGACG 463
 Db 123 GGCTACTACTCTCTTCT 173
 QY 464 GTCAAGTGGGTGGGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 523
 Db 174 CTGGGGTGGTACAGGCTGAACCGTCTACCCCACTAGCAGGCAACCTGGCCACGCG 233
 QY 524 GCTATTCGATGGACCTCGCGATTTTGGCGGTTCACCTTGGGTGGCTCTCTCGATCAGG 583
 Db 234 GAGCTTCGGTGGCTTAACCATCTTCTCTCCACCTGGCGCGTCTCTCATCGATTTAG 293
 QY 584 GCGGATCAACATGATCAGCACTTCTTGAACATGCGCGCCCGGCGATGACGCTGCA 643
 Db 294 GGGCCATTAATTTTACCAACCATCATTAACATAAACCACGCACTCTCCCAATATC 353
 QY 644 AAGTGGCGTGGTCTCGTGTGATCTTTATCAGGCTTGGGTGATCTCTGCTGGCGCTGC 703
 Db 354 AGACCCCTCTCTTCTGATGAGCGCTCTAATTAAGTGGCTCTCTCTCTCTCTCTCTCT 413
 QY 704 CGGTCTCGGTGGTGAATCACCATCTCTGACCGACCGCTAATCTGGCAGACCTTCT 763
 Db 414 CAGTCTCGCAGCGGATCAATGCTTCTAAGACCGAAACCTAAACACCACTTTCT 473
 QY 764 TCAATCTGCTGGCGGGTGAACCCGATTTCTGTACCAACACATCTCTGTGGTCTTTGGGC 823
 Db 474 TCGACCCCTCGAGCGGAGGAGACCCAACTTTTACCACTATTTCTGATTTCTTTGGCC 533
 QY 824 ACCGGAAGTGTACATCATCTCTGCGGCTTTGGCATCATCAGCACTGCTGTGCA 883
 Db 534 ACCAGAAGTGTACATCTCTCACTTACCAGGCTTTGGAATAATCTCCCATATTTGTGCT 593
 QY 884 CCT---TCTGAAAAAGCGGCTTCTCGTATCTGCGGATGGTCTATGCAATGGTGGCAA 940
 Db 594 ACTATCGCGCAAAAAAGAACCACTTCGGCTACATGGCATGTTTGGGCCATGATGCTTA 653
 QY 941 TCGGTGTTCTGGGCTTTGCTGCTGGCGCGACCAATGATACACCGTGTGTGTGCTGCTGA 1000
 Db 654 TTGGCCCTTCTAGGCTTCATCGTCTGAGCCCACTATGTTTACGTTAGGAATGGACGTTG 713
 QY 1001 CCAGCAATCTCTACTTCACTGCTGGCCACCATGGTGTATCGCGTGGCCGCACTAAGA 1060
 Db 714 ACACCGCGCTTACTTCACTCGCCACATAATTTATGCGCATCCCAACCGAGTAAAG 773

1061 TCTTCTCGTGATGCCACGATGGGGCGCTCGGTTGAGTTCAAAATGCCGATGCTCT 1120
 Db TATTAGCTGACTAGCCACCCCTTCAACGGGGCTCCATTAATGAGAAATCCAAATGCTCT 833
 1121 GGGGCTTTGGCTTATGTTCTGTTTACCGTGGGTGGTGTGACCGGTATGCTGTCGCCC 1180
 Db GAGCCCTAGGCTTCTCTCTTACCGTAGGCGGACTAACAGGCAATGCTTATGCCA 893
 1181 AAGCGGGTCTGGACGCTGATACACGACACTATTAGTGGTGGCGCACTTCCATTATG 1240
 Db ACTCTCTCTAGAGTAGTCTGATGACACTACTAGTAGTTGCCCACTTCCACTACG 953
 1241 TGATGTCGCTGGGTGCGAATTTGCGATTTTCGCGGTATCTATTTTACATGCGGAAT 1300
 Db TCCTATCAATAGGACCGCTATTGTCATCATGGGTGCTTTTGTACACTGATTTCCACTT 1013
 1301 TCTGGGCGGCTTTCCCGGAATGGGCTGCAAGCTGCACTTCTGGACCTTCTTCAATG 1360
 Db TCTAGGATACACCCCTCCACAGCCTGAACAAATCCATTTTGGAGTAGTTCTGTAG 1073
 1361 GTGCGAAGCTCACGTTTCCCGCAGCACTTCTGGGACGTCAGGATGCGCGCGGTT 1420
 Db GGGTTAACCTTAACCTTCTTCCCAACACTTCTCGGACTCGCGGAATGCGCGTGGAT 1133
 1421 ACATGCACTATCCGAGCCTTCGCGTGTGGACAAAGTCTGCTCTATGTTGGTTC 1480
 Db ACTCAGATTACCGACGCTACAGCTATGAACACAGTTTCTCCATTGGCTCTCTAA 1193
 1481 TGGCCTTCGCTCGTTCCTGTTCTTCACTGATCTT 1517
 Db TCTCATTATCGCTGCTATATTTCTTATTTATCTT 1230

RESULT 10
 BC024102
 LOCUS
 DEFINITION Homo sapiens, clone IMAGE:3508235, mRNA.
 ACCESSION BC024102
 VERSION BC024102.1 GI:18848186
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 3167)
 Strausberg,R.
 Direct Submission
 Submitted (19-FEB-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILNL)
 DNA Sequencing by: Genome Sequence Centre,
 BC Cancer Agency, Vancouver, BC, Canada
 info@bgsc.bc.ca
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
 Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
 Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
 Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stott,
 Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
 George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LILNL at: <http://image.llnl.gov>
 Series: IRAL Plate: 23 Row: e Column: 21

This clone has the following problem: frame shifted.

FEATURES

Location/Qualifiers
 1..3167

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3508235"
 /issue_type="Brain, neuroblastoma"
 /clone_lib="NIH MGC_19"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"

ORIGIN

Query Match 23.7%; Score 397.2; DB 3; Length 3167;
 Best Local Similarity 58.1%; Pred. No. 1.9e-95;
 Matches 743; Conservative 0; Mismatches 523; Indels 12; Gaps 2;

244 AACGGACACCTGGGAACGTCATGGTTACCTACCATGTATTCTGATGATGTTCTTTGTG 303
 Db AACGACCACATCTACACGTTATCGTCACAGCCATGCAATTTGTAATAATCTCTTTATA 1756
 304 GGTATCCCGCATGTTCCGTGGTTTGTGTAATATCTGATGCCGTGCAAAATCGGCGGT 363
 Db GTAATACCCATCAATAATCGAGGCTTTGGCACTGACTAGTTCCCTTAATAATTCCTTTATA 1756
 364 CCGGATATGGCCTTCCCGGGTATGAACAACTGTCTGTTCTGGCTGTTCTATTCGCGGTACC 423
 Db CCGGATATGGCCTTCCCGGGTATGAACAACTGTCTGTTCTGGCTGTTCTATTCGCGGTACC 423
 424 GCGATGGGCGTGGCTGCTGCTGTTCCGACCGGGCGGTGACGGTCAGTGGTGGGCGGTT 483
 Db CTACTCTGCTGCATCTGCTATAGTGGAGCGCGGAGCAGGAACAGGTTGAACAGTCT 1933
 484 GGTGGGTTCTGTACCGCGCTGTGACCGCGGAACTACTCTCCACCTGGAGCCTCCGTAGACCTAAC 543
 Db ATTTTCGCGGTTCACATGTTGGGTGCTCTCTGATCATGGGCGCGATCAACATGATCAAG 603
 1988 ATCTTCTCCTTACACCTAGCAGGTGCTCTCTCTATCTTAGGGGCCATCAATTTCAATCACA 2047
 604 ACCTTCTTGAACATGGCGCGCGCGCATGACGCTGCACAAAGTGGCTGTTCTCGTGG 663
 Db ACAATTATCAATATAAAACCCCTGCGATTAACCAATACCAAGCGCCCTCTTCGCTCTGA 2107
 664 TCGATCTTTATACGGGCTTGGCTGATCTGCTGGCGGTGCGGTTCTGGCTGGTGGCAATC 723
 Db TCGTCTTATACAGAGTCTTCT 2167
 724 ACAATGCTGTGACCGGACCGTAACTTGGGACGACCTTCTTCAATCTGCTGGCGCGGT 783
 Db ACTATACTACTAACAGACCGCAACCTCAACACCACCTTCTTCGACCCCGCGGAGGAGGA 2227
 784 GACCCGATTCTGTACCAACACATCTCTGTTCTTTGGGACCCCGGAAGTGTACATCATC 843
 Db GACCCGATTCTGTACCAACACATCTCTGTTCTTTGGGACCCCGGAAGTGTACATCATC 843
 844 ATTTCCGCGGCTTTGGCATCATCAGCCATGTCGTGTGACCTTCTC 900
 Db ATTTCCGCGGCTTTGGCATCATCAGCCATGTCGTGTGACCTTCTC 900
 2288 ATCTTACAGGCTTCGGATAAATCTCCCAATTTGTAATTTACTACTCCGGAAGAAAGAA 2347
 901 GTCTTCGGTTACCTGCGGATGTTATGCAATGGTGGCAATCGGTGTTCTGGGCTTTGTC 960
 Db CCATTTGGATACATAGGTATGGTCTGAGCTATGATATCAATTTGGCTTCTTAGGTTTATC 2407
 961 GTCTGGGCGGACCATGATGATACCGTTGGTATGTCGTGACCCAGCAATCCTTACTTCAATG 1020
 Db GTGTGACACACCATATATTACAGTAGGAATAGACGTAGACACACAGCATATTTTCAACC 2467
 1021 CTGGCCACCATGTTGATGCGGTCGCCGCGGATTAAGATCTTCTCTGTTGATCGCCACG 1080
 Db TCCGCTACCAATATCATCGTATCTCCACCGCGGTCAAGTAATTTAGCTGACTGCCACA 2527

QY 1081 ATGTGGGGCGCTCGGTGAGTTCAAAATCGCCGATGCTCTGGGCTTTGGCTTTATGTTTC 1140
 Db 2528 CTCACCGAAGCAATATGAATGATCTGCTGCACTGCTCTGAGCCCTAGGATTCATCTTT 2587
 QY 1141 CTGTTACCGTGGGTGGTGTGACCGGTATCGTGTGCGCCCAAGCGGTCTGGACCGTGCA 1200
 Db 2588 CTGTTACCGTGGGTGGTGTGACCGGTATCGTGTGCGCCCAAGCGGTCTGGACCGTGCA 1267
 QY 1301 TATCAGACACCTATTACGTGGTGGCGCACTTCATATATGATGTCGTGGTGGCGATC 1260
 Db 2648 CTACACACACGCTACTAGTGTGAGTCACTTCCACTATGCTCCATGATCTCCATATCAATAGAGCTGTA 2707
 QY 1261 TTTGGCATCTCGCGGTATCTACTTTTACATGCCGAAGTTCTCGGCGCGCTTTCCCG 1320
 Db 2708 TTTGGCATCATAGAGGCTTCACTCACTGATTTCCCTATTCTCAGGCTACACCTAGAC 2767
 QY 1321 GAATGGGCTGGAAGCTGCACTTCTGACCTTCTGACCTTCTCATGGTGGCAAGCTCACTGTTCTTC 1380
 Db 2768 CAACCTTACGCCAAATCCATTTCACTATCATATTCATCGGCGTAAATCTAACTTTCTTC 2827
 QY 1381 CCGCAGCACTTCTGGGACGTCAGGATATGCGCGCGGTATACATCGACTATCCGAGCC 1440
 Db 2828 CCACACACCTTCTGGCCCTATCGGAATGCCCGGAGTTACTCGGACTACCCCGATGA 2887
 QY 1441 TTCCGCTGTGGACAAAGTCTCGTCTCTATGTTGCGTTCCTGGCCTTCGCTCGTTCTCTG 1500
 Db 2888 TACACACATGAACATCTCTATCATCTGAGGCTCATTTCTCTTAACAGCAGTAATA 2947
 QY 1501 TTCTTCATCTGATCTTT 1518
 Db 2948 TTAATAATTTTCAGATT 2965

RESULT 11
 CR597319
 LOCUS
 DEFINITION full-length cDNA clone CS0DE005YP22 of Placenta of Homo sapiens (human).
 ACCESSION CR597319
 VERSION 1
 KEYWORDS HTC; CNSLT; cDNA.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished
 CONTACT : Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue
 Redwood City, CA 94063
 Genoscope.
 Direct Submission
 JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 COMMENT - Web : www.genoscope.cns.fr)
 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies, a division of Invitrogen.

FEATURES

source
 1..1486
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DE005YP22"
 /tissue_type="placenta"
 /plasmid="pCMVSPORT_6"

ORIGIN

Query Match 23.7%; Score 396.2; DB 3; Length 1486;
 Best Local Similarity 58.1%; Pred. No. 2.9e-95;
 Matches 743; Conservative 0; Mismatches 523; Indels 13; Gaps 2;
 QY 244 AACGACACCTGTGGAACTCATGTTACCTTACCATGGTATCTGATGATGATCTTTTGTG 303
 Db 140 AACGACACCTGTGGAACTCATGTTACCTTACCATGGTATCTGATGATGATCTTTTGTG 199
 QY 304 GGTATCCCGCATTTGTTGGTGGTATCTGATGATGATGATCTTTTGTG 363
 Db 200 GTAATACCATATAATCGAGGCTTTGGCACTAGTATGCTTCCCTAATAATCGGTGCC 259
 QY 364 CGGATATGGCTTTCCCGCATGATGAACAACTGTCGTTCTGGTGTTCATTCGGGTACC 423
 Db 260 CCGATATGGGCTTTCCCGCATGATGAACAACTGTCGTTCTGGTGTTCATTCGGGTACC 319
 QY 424 GCGATGGGCGTGGCTTGGCTTTCCGACCGGCGGTGACGCTCAGCTGGGTTCGGGCGTT 483
 Db 320 CTACTCTCTGCTTGCATCTGCTATAGTGGAGCGCGCAGGAACAGGTGGAACAGTC--- 376
 QY 484 GGTGGGTTCTGTACCGCGCTGTGACCGCGGAAGCTGCTATTCGATGGACCTCGG 543
 Db 377 -----TACCTCTCTGCTGGAGGAATCTCCACACCTCGAGCTCCGTAGACCTAAC 430
 QY 544 ATTTTCGCGTTCATTTGTCGGTGCCTCTCGATCATGGCGGATCAACATGATCAACG 603
 Db 431 ATCTTCTCTTACACTAGCAGGTATCTCTCTATCTTAGAGCCATCAATTTTCATACA 490
 QY 604 ACCTTCTTGAACATGCGCGCGCGCATGACGCTGCACAAAGTGGCTGTTCTCGTGG 663
 Db 491 ACAATTATTAATAAACCCTGCTGCAATACCAACATACCAACAGCCCTTTTCGTGA 550
 QY 664 TCGATCTTTTATCAGGCTTGGCTGATCTCTGCGGCTGCGGCTTCTGGTGTGCAATC 723
 Db 551 TCCGCTCTATCAGCAGCTTACTTCTCTCTCTCTCCAGTCTTAGCGCTGGCATC 610
 QY 724 ACCATGCTGTGACCGACCTAACTTCGGCAGCACTTCTTCAATCTGCTGGCGGCT 783
 Db 611 ACTATACTACTAACAGACCTTAACCTCAACACACCTTCTTCGACCCAGCGGAGGAG 670
 QY 784 GACCGATCTGTACCAACACATCTCTGTTCTTTGGGACCGGAGTGTACATCATC 843
 Db 671 GACCCCATCTTATACCAACACCTTCTGATTTTTCGGTCACTGATGATCAATTTGGCTTCTAGGTTAT 730
 QY 844 ATTTGCTCCGCTTTGGCATCATCAGCCTGCTGCTGCGACCTTCTC---GAAAAGCC 899
 Db 731 ATCTTACAGGCTTCGGAATACTCCCATATTGTAATTTACTTACTCCGGGAAAAAGA 790
 QY 900 GGTCTTCGGTACCTGCGGATGCTATGCAATGGTGGCAATCGGTGTTCTGGGCTTTGT 959
 Db 791 ACCATTTGGATACATAGGTATGCTGATGATGATGATGATGATGATGATGATGATGAT 850
 QY 960 CGTCTGGGCGCACCATGTACACCGTGGTATGTCGCTGACCCAGCAATCTTACTTCAT 1019
 Db 851 CGTGTGACACACCATATAATTACAGTAGGAATAGAGTAGACACACAGGACATATTTCAC 910
 QY 1020 GCTGGCCACCATGGTATCGCGTGGCGGACCGCATTAAGATCTTCTGTGGATTCGCCAC 1079
 Db 911 CTCGGTACCATTAATCATGCTATCCCGACCGCGCTCAAAGTATTTAGTGTACTCGCCAC 970
 QY 1080 GATGTGGGCGGCTCGGTTGAGTTCAATTCGCGATGCTCTGGGCTTTGGCTTTATGTT 1139
 Db 971 ACTCCACGGAAGCAATATGAAATGATCTGCTGAGTGCTCTGAGCCCTAGGATTTATTT 1030
 QY 1140 CTGTTTACCGTGGTGGTGTGACCGGTATCGTGTGGCCCAAGCGGTCTCGACCGTGC 1199
 Db 1031 TCTTTTACCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1090
 QY 1200 ATATCAGCACCTATTACGTGGTGGGACCTTCCATTTATGATGCTCGGTGGTGGAT 1259
 Db 1091 ACTACAGCACGCTACTAGTGTAGCCCATCTTCCATTTCCCTATGCTCTATCAATAGGAGCTGT 1150
 QY 1260 CTTTGGCATCTTCGCGGTATCTATTTTACATGCGCAAGTTCCTGGGCGCGCTTCC 1319


```
Db 1151 ATTTGGCATATAGAGGCTTCATCTAGTATTTCCCTATTTCTCAGGCTACACCTAGA 1210
Qy 1320 GGAATGGGCTGCAAGCTGCACTTCTGGACCTTTCTTATCGGTGGGAACGTCAGTTCTT 1379
Db 1211 CCAAACTTACGCCAAATCCCAATTTGGTATCATATTCATCGCGCTAAATCTAATCTTCTT 1270
Qy 1380 CCGCAGACACTTCTTGGGACGTCAGGGTATCCGCGCGGTTTACATCGACTATCCCGAAGC 1439
Db 1271 CCACAAACACTTCTCGGCTTATCGGAATGCCCGAGCTTACTCGGACTATCCGATGC 1330
Qy 1440 CTTGCGGCTGTGGAAACAAGTCTGTCTATGTGCGTCTGCTGCGCTTCGCTGCTTCTT 1499
Db 1331 ATACACCAATGAATATCTATCATCTGTAGGCTCATCTCAATTTCTCTAACAGAGTAAT 1390
Qy 1500 GTTCTTCATCGTATCTTT 1518
Db 1391 ATTAATAATTTTCATAAT 1409

RESULT 12
CR590798 1499 bp mRNA linear HTC 21-JUL-2004
LOCUS full-length cDNA clone CS0DC025YC06 of Neuroblastoma Cot
DEFINITION 25-normalized of Homo sapiens (human).
ACCESSION CR590798
VERSION CR590798.1 GI:50471605
KEYWORDS HTC; CNSLT cDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1499)
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue
2 (bases 1 to 1499)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
source
Location/Qualifiers
1..1499
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC025YC06"
/tissue_type="Neuroblastoma Cot 25-normalized"
/plasmid="pCMVSPORT_6"

ORIGIN
Query Match 23.7%; Score 396.2; DB 3; Length 1499;
Best Local Similarity 58.1%; Pred. No. 2.9e-95;
Matches 743; Conservative 0; Mismatches 523; Indels 13; Gaps 2;

Qy 244 AAGGACACTGTGGAACTGATGTTACTACCATGGTATTCGATGATGTTCTTTTG 303
Db 141 AAGACACATCTACAACTGTTTCGTCAGCCCATGCAATTTGTAATAATCTTTCATA 200
Qy 304 GGATCCCCGCAATGTTGGTGGTGTGTTTGGTAATATCTGATCGCGCTGCAATTCGCGCT 363
Db 201 GTAATACCATATAATCGGAGCTTTGGCACTAGTCTCCCTAATAATACGTTGCC 260
Qy 364 CCGATATGSCCTTCCCGGTATGACAACTTCGTTCTGGTGTTCATTGCCGTACC 423
```

```
Db 261 CCGATATGCGGTTTCCCGCATAAACAATAAGCTTCTGACTTCTACCCCTCTCTC 320
Qy 424 GCGATGGGCTGGCTTCGCTTCGCTGTTGCGACCGGGGTGACGGTCACTGGGTTCGGGCGTT 483
Db 321 CTACTCTCTCTTGCACTCTGCTATAGTGGAGGCGCGGAGGAACAGGTTGAACAGTC--- 377
Qy 484 GGTGGGTTCTGTATCCCGCGCTGTGCGACCGGGAAGCTGGCTATTCGATGGAACCTCGG 543
Db 378 -----TACCCTCCCTTGGCAGGAACTACTCCACCTTGGAGCCCTCCGTAGACCTAAAC 431
Qy 544 ATTTTCGCGGTTTCACTTGTGGGTGCTCTCGATCATGGGGCGGATCAACATGATCAAG 603
Db 432 ATCTTCTCTTACACTAGAGTATCTCTCTATCTTAGAGGCATCAATTTCAACACA 491
Qy 604 ACCTTCTTGAACATGCGCGCCCGCGCATGACCTGCAACAAAGTGCCTGTTTCTCGTG 663
Db 492 ACAATTATTAATAAATAAACCCTGCTAATACCAATACCAACAGCCCTTTTCGTCGTA 551
Qy 664 TCGATCTTTATCAGGCTTGGCTGATCCTGCTGGCGCTGCGGTTCTGGCTGTGCAATC 723
Db 552 TCCGTCCTAATACAGAGTCTTACTTCTCTATCTCTCCAGTCTTAGCCGTGCAATC 611
Qy 724 ACCATGCTGTGACCGACCGTAATCTTGGCACGACCTTCTTCAATCTCTGTCGCGCGGT 783
Db 612 ACTATACTACTAACAGACCGTAACCTCAACACCACCTTCTTCGACCGCGGAGGAGA 671
Qy 784 GACCGATCTGTACCAACACATCTCTGTTGTTTGGCACCGGGAAGTGTACATCATC 843
Db 672 GACCCCAATCTATACCAACACCTTATCTGAATTTTCGGTCACTCAATTTTATATTTCTC 731
Qy 844 ATTCTGCCCGCTTGGCATCATCAGCCATGTCGTGTCGACCTTCTC-----GAAAAGCC 899
Db 732 ATCTTACGAGCTTCGGAATAATCTCCATATTTGTAATCTTACTTACTCCGGGAAAAAAGA 791
Qy 900 GGTCTTCGGTTACTGCGGATGCTATGCAATGGTGGCAATCGGTGTTCTGGGCTTTGT 959
Db 792 ACCATTTGGATACATAGTATGCTGAGCTATGATATCAATTTGGCTTCTCGGGTTTAT 851
Qy 960 CGTCTGGGCGCACCATGATACACCGTGTGATGTCGTGACCCAGCAATCTTCTTCTCAT 1019
Db 852 CGTGTGAGCACCATATATTTACGTAGAAATAGAGTAGACACAGCATATTTTCAAC 911
Qy 1020 GCTGGCCACCATGTTGATCGCGGTGCGACCGGATTAAGATCTTCTGTGGATCGCCAC 1079
Db 912 CTCGCTTACCATATATCATCTGATCCCAACCGGCTCAAAAGTATTTAGCTGACTCGCCAC 971
Qy 1080 GATGCGGGGCTCGGTTGAGTTCAATCGCGATGCTCTGGGCTTTTGGCTTTATGTT 1139
Db 972 ACTCCACGGAAGCAATGAAATGATCTGCTGAGTCTCTGAGCCCTAGATTTATTTT 1031
Qy 1140 CTTTTCACCGTGGGTGTTGACCGGTATCGTGTGCCCCAACGCGGTCTGGACCGTGC 1199
Db 1032 TCTTTTACCGTAGTGGCTGACTGCAATTTAGCAAACTCATCACTAGACATCGT 1091
Qy 1200 ATATCAGACACCTATTAAGTGTGGCGACTTCCATATGTAATGCTGCTGGTGGTGGAT 1259
Db 1092 ACTACAGCACGTAATACGTTGAGCCCACTTCCACTATGTCCTATCAATAGAGTGT 1151
Qy 1260 CTTTGGCATCTTCGCGGTATCTACTTTTACATCCGAAGTTCTCGGCGCGCGCTTTCCC 1319
Db 1152 ATTTGGCATATAGAGGCTTCACTAGTATTTCCCTATTTCTCAGGCTACACCTAGA 1211
Qy 1320 GGAATGGGCTGCAAAAGCTGCACTTCTGGACCTTCTTATCGGTGGGAACGTCAGTTCTT 1379
Db 1212 CCAAACTTACGCCAAATCCCAATTTGGTATCATATTCATTCGCGCTAAATCTAATCTTCTT 1271
Qy 1380 CCGCAGACACTTCTGGAAGCTGAGGTATGCGCGCGGTGATACATGCACTATCCCGAAGC 1439
Db 1272 CCCAACACACTTCTCGGCTATCCGGAATGCCCGAGTCTTACGAGCTATCCCGATGC 1331
Qy 1440 CTTGCGGCTGTGGAACAAAGTCTCGTCTCTATGTTGCGGTTCTGGCTTCGCTGCTCT 1499
```


CR615203.1 GI:50496010
HTC; Cnslt cdna.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1504)
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished
Contact : Peng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
Faraday Avenue
2 (bases 1 to 1504)
Genoscope.
Direct Submission
Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
1st strand cDNA was primed with a NotI-cligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with NotI and cloned
into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES
Location/Qualifiers
1..1504
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF025YA19"
/tissue_type="Fetal brain"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 23.7%; Score 396.2; DB 3; Length 1504;
Best Local Similarity 58.1%; Pred.No.2.9e-95;
Matches 743; Conservative 0; Mismatches 523; Indels 13; Gaps 2;
QY 244 AACGACACCTGTGAACCTCATGTGTACTACCATGTGTATCTCATGATGTCTTTGTG 303
DB 86 AACGACACATCTACACAGTATATCGTCACAGCCCATGATTTGTAATATCTTTCATA 145
QY 304 GGTATCCCGCATGTGTGGTGGTTTGTAACTATCTGATCCGCTGCAATCGCGGT 363
DB 146 GTAATACCATCATATATCGAGGCTTTGGCAACTGACTAGTTCCCTTAATATCGGTGCC 205
QY 364 CCGGATATGCTTCCCGGTATGAACACCTGTGTTCTGCTGTTTATTCGCCGTACC 423
DB 206 CCGATATGCGGTTTCCCGCATTAACACATAGCTTCTGACTCTTACCCCTCTCTC 265
QY 424 GCGATGGGCGTGGCTTGGCTGTTCGCACCGGGCGGTGACGGTCAGTGGGTTTCGGCGGT 483
DB 266 CTACTCCTGCTGCTATCTGCTATAGTGGAGCGCGGCGGAGGTTGAACAGTC--- 322
QY 484 GGTGGGTTCTGTACCCGCGGTGTCGACCCGGAGCTGGCTATTCGATGACCTCGG 543
DB 323 -----TACCCCTCCCTTGGCAGGAATACTCCACCCCTGGAGCTTCGTAGACCTAAC 376
QY 544 ATTTTCGCGGTTCACTTGTGCGGTGCTCTCTGATCATGGGCGGATCAACATGATCAG 603
DB 377 ATCTTCTCTTACACCTAGCAGGTATCTCTCTATCTTAGGAGCCATCAATTCATCACA 436
QY 604 ACTTCTTGAACATGGCGGCCCCCGGATGACGCTGCACAAAGTGCGGTTGTTCGTGG 663
DB 437 ACAATTATTAAATAAAACCCCTGGCATAACCCAAATACCAACGCCCTTTTCGTCTGA 496
QY 664 TCGATCTTTATACCGGTTTGGCTGTATCTCTGCTGGCGTCTGGCTGTGGTGAATC 723
DB 497 TCGTCTCTATACAGAGTCTTACTTCTCTATCTCTCCAGTCTTAGCCGCTGGCATC 556
QY 724 ACCATGCTGTGACCGACCGTAACTTTCGGCAGCACTTCTTCAATCTCTGCTGGCGCGT 783
DB 557 ACTATATACTAACAGACCGTAACTTCAACACCACTCTTTCGACCCAGCGGAGGAGA 616

QY 784 GACCCGATTTCTGTACCAACACATCTCTGTGGTCTTTGGGACCCCGGAAGTGTATCATC 843
DB 617 GACCCGATTTCTGTACCAACACATCTCTGTGGTCTTTGGGACCCCGGAAGTGTATCATC 676
QY 844 ATTCTCGCGGCTTTGGGATCATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 899
DB 677 ATCTACACGCTTTGGGATTAATCTCCATATTTGTAATCTTACTCTCCGGGAAAAAAGA 736
QY 900 GGTCTTCGGTTACTCGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 959
DB 737 ACCATTGGATACATAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 796
QY 960 CGTCTGGGCGACACATGTACACCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1019
DB 797 CGTGTGAGCACACATATATTACAGTAGAATAGACGTAGACACACAGGACATATTTCAC 856
QY 1020 GCTGGGACCATGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1079
DB 857 CTCGGCTACCATTAATCATCTGCTATCCCGCGCTCAAGATATTTAGCTGACTCGCCAC 916
QY 1080 GATGTGGGCGGCTCGGTTGAGTTCAAAATCGCGATGCTCTGGGCTTTGGCTTTATGTT 1139
DB 917 ACTCCAGGAACAATAGAAATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 976
QY 1140 CCTGTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1199
DB 977 TCTTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1036
QY 1200 ATATCAGCACATTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1259
DB 1037 ACTACAGCACGCTACTACGTTGAGCCCACTTCCACTTATCTGCTGCTGCTGCTGCTGCT 1096
QY 1260 CTTTGGCATCTTCCGCGTATCTATTTTACATGCGCAAGTTCTCGGGCGCGCTTTTCCC 1319
DB 1097 ATTGGCATCATAGGAGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1156
QY 1320 GGAATGGCTGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1379
DB 1157 CCAACCTCAGCCAAAATCCATTTGCTATCATATTCATCGGGTAAATCTAACTTTCTT 1216
QY 1380 CCGCAGCAGCTTCTCGGAGCTCAGGGTATCGCGGCTTATCATGCTGCTGCTGCTGCTGCT 1439
DB 1217 CCCACACATTTCTCGGCTATCCGGAATGCCCGAGCTTACTCGGACTATCCGATGC 1276
QY 1440 CTTCCGCTGTGGAACAAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1499
DB 1277 ATACACACATGAAATATCTTATCATCTGAGGCTCATTTCTTCTAAACAGCAGTAAT 1336
QY 1500 GTTCTTTCATCGTGTCTTT 1518
DB 1337 ATTAATAATTTTATAATT 1355

RESULT 15
CR621733
LOCUS
DEFINITION full-length cDNA clone CS0CAP002y018 of Thymus of Homo sapiens (human).
ACCESSION CR621733
VERSION CR621733.1 GI:50504540
KEYWORDS HTC; Cnslt cdna.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1510)
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished
Contact : Peng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
Faraday Avenue

REFERENCE 2 (bases 1 to 1510)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-strand cDNA was digested with Not I and cloned
into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
was normalized. Library was constructed by Life Technologies, a
division of Invitrogen.
FEATURES Location/Qualifiers
source 1..1510
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clones="CS0CAP002Y018"
/tissue types="Thymus"
/plasmid="pCMVSPORT_6"
ORIGIN
Query Match 23.7%; Score 396.2; DB 3; Length 1510;
Best Local Similarity 58.1%; Pred. No. 2.9e-95;
Matches 743; Conservative 0; Mismatches 523; Indels 13; Gaps 2;
QY 244 AACGGACACCTGTGGAACGTCATGGTTACCTACCATGGTATCTCTGATGATGTTCTTTGTG 303
DB |||||
QY 77 AACGACCAATCTACAAAGTTATCGTCAGAGCCATGCAATTTGTAATATCTTTCTTCATA 136
DB |||||
QY 304 GGTATCCCCGCAATGTTCCGTGGTGTGTTGTAATCTATCTGATGCCGTGCAAAATCGCGCT 363
DB |||||
QY 137 GTAATACCCATCAATATCGGAGGCTTTGGCACTGACTAGTTCCTCCCTAATAATATCGGTGC 196
DB |||||
QY 364 CCGGATATGCTTCCCGCTATGACAACTCTGCTGCTGGGTGTTCTATTGCGGTAC 423
DB |||||
QY 197 CCGGATATGCTTCCCGCTATGACAACTCTGCTGCTGGGTGTTCTATTGCGGTAC 256
DB |||||
QY 424 CGGATGGCGTGGCTTGGCTGTTGCGACCGGGGGTGAAGGTGAGTGGGTTCGGGCGTT 483
DB |||||
QY 257 CTACTCTGCTTGCATCTGCTATAGTGAGGCGCGGAGGACAGGTTGAACAGTC--- 313
DB |||||
QY 484 GGTGGGTTCTGTACCCGCGCTGTGACCCGGAAGCTGGCTATTCGATGACCTCGG 543
DB |||||
QY 314 -----TACCCCTCCCTTGGCAGGAACTACTCCACCCCTGGAGCCCTCCGTAGACCTAAC 367
DB |||||
QY 544 ATTTTCGGGTTTCACTGTGCGGTGCTCTCTGATCATGGGCGCGATCAACATGATCAG 603
DB |||||
QY 368 ATCTTCTCTTACACCTAGCAGTATCTCTCTATCTTAGGAGCCATCAATTTTCATCA 427
DB |||||
QY 604 ACCTTCTTGAAATGCGGCGCGCGGATGACGCTGCAAAAGTGCCTGTTGTTCTCGTGG 663
DB |||||
QY 428 ACAATTATTAATAAAACCCCTGCGCATAAACCAATACCAACGCGCCCTTTTCGTCTGA 487
DB |||||
QY 664 TCGATCTTTATCAGGCTTGGGTGATCTCTGCTGGCGCTGCGGTCTCTGCTGTCGATC 723
DB |||||
QY 488 TCGGTCTTAATCAGAGGCTTACTTCTCTATCTCTCCAGTCTAGCCGCTGGCATC 547
DB |||||
QY 724 ACCATGCTGTCGACCGAATGATCTTGGGACGACCTTCTTCAATCCTGCTGGCGCGGT 783
DB |||||
QY 548 ACTATACTACTAACAGACCGTAACTCTCAACACACCTTCTTCCAGCCCGGAGGAGGA 607
DB |||||
QY 784 GACCCGATCTGTACCAACACATCTGTTGTTTGGGACCCCGGAGGTGTACATCATC 843
DB |||||
QY 608 GACCCGATCTGTACCAACACATCTGTTGTTTGGGACCCCGGAGGTGTATTTCTC 667
DB |||||
QY 844 ATTTCTGCGCGCTTTGGGATCATCAGCCATGTCGTGTCGACCTTCTC-----GAAAAAGCC 899
DB |||||
QY 668 ATCTTACAGGCTTCGGAATATCTCCCATATTTGTAATTTACTACTCTCGGGAAAAAAGA 727
DB |||||
QY 900 GGTCTTCGTTACTGCGGATGCTCTATGCAATGGTGGCAATCGGTGTTCTCGGCTTGT 959
DB |||||
QY 728 ACCATTGGGATACATAGGTATGCTGAGCTATGATCAATTTGGCTTCTTAGGCTTTAT 787
DB |||||

QY 960 CGTCTGGCGCACCAATGTTACACGGTTGGTATGTGTGCTGACCCAGCAATCCTACTTCAAT 1019
DB |||||
QY 788 CGTGTGAGCACACCATATATTTACAGTAGGAATAGACGTAGACACAGGACATATTTAC 847
DB |||||
QY 1020 GCTGGCCACCATGTTGATCGGGTCCGACCGGATTAAGATCTTCTCGTGATCGGCAC 1079
DB |||||
QY 848 CTCGCTACCAATAATCATCGCTATCCCGACCGGCTCAAGTATTTAGCTACTCGGCAC 907
DB |||||
QY 1080 GATGTGGGGCGGCTCGGTGAGTTCAAAATCGCCGATGCTCTGGGCTTTGGCTTTATGTT 1139
DB |||||
QY 908 ACTCCACGGAAGCAATATGAAATGATCTGTCAGTGTCTGAGCCCTAGGATTTATTT 967
DB |||||
QY 1140 CTTGTTTACCGGGTGGTGTGACCGGTATGCTGTGCCCCAAGCGGTCTGGACCCGTC 1199
DB |||||
QY 968 TCTTTTACCGGTAGTGGCTGACTGGCAATTTATTAGCAAACTCATCACTAGACATCGT 1027
DB |||||
QY 1200 ATATCAGCACACCTATTACGTGGTGGGCGCACTTCCATTTATGTGATGTCGCTGGGTGGAT 1259
DB |||||
QY 1028 ACTACACGACAGTACTAGTGTGAGCCACTTCCACTATGCTCTATCAATAGGAGCTGT 1087
DB |||||
QY 1260 CTTTGGGATCTTCGGCGTATCTACTTTTATCATGCCGAAGTTCTCGGCGCCGCTTTCCC 1319
DB |||||
QY 1088 ATTTGCCATCATAGGAGGCTTCAATCACTGATTTTCCCTATTTCTCAGGCTACACCTAGA 1147
DB |||||
QY 1320 GGAATGGGCTGCAAGGCTGCACTTCTGGACCTTCTTATCGGTGCGAAGCTCAAGTTCTT 1379
DB |||||
QY 1148 CCAACCTACGCCAAATCCATTTGCTATCATATTCGCGGTAAATCTAACTTTCTT 1207
DB |||||
QY 1380 CCGCAGCACTTCTCGGAGCTGAGGTATGCGCGCGCTTACATCGACTATCCCGAAGC 1439
DB |||||
QY 1208 CCCAACACACTTCTCGGCTTATCCGGAATGCCCGGACGTTACTCGGACTATCCCGATGC 1267
DB |||||
QY 1440 CTTGCGGTGGAACAAAGCTCTGCTCTATGTTGGTGGTTCCTGCGCTTCTCGCTCGTTCT 1499
DB |||||
QY 1268 ATACCAACATGAATATCCCTATCATCTGAGGCTCAATTTCTTAAACAGAGTAAT 1327
DB |||||
QY 1500 GTTCTTTCATCGTATCTTT 1518
DB |||||
QY 1328 ATTAATAATTTTCATAAT 1346
DB |||||

Search completed: October 22, 2004, 18:59:14
Job time : 7641 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 23:27:21 ; Search time 157 Seconds
(without alignments)
1272.689 Million cell updates/sec

Title: US-09-712-768a-2
Perfect score: 3000
Sequence: 1 MADAAIHGHHEKQGFTR.....PAHTFTLPKRSDWDKHPSH 557

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues
Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_23Sep04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3000	100.0	557	4	AAY97750 G. oxydan
2	1543	51.4	514	8	ADJ79757 Subunit I
3	1542	51.4	514	7	AD661901 Rat Prote
4	1531	51.0	513	6	ABR64214 Angiogene
5	1531	51.0	513	8	ADJ35456 Human cyr
6	1528	50.9	513	7	ADJ68385 Human hea
7	1469.5	49.0	530	6	ABJ18768 Pseudomon
8	1278	42.6	425	7	ABO68563 Pseudomon
9	1223	40.8	573	6	ABU336908 Protein e
10	1223	40.8	573	6	ABU34934 Protein e
11	1215	40.5	574	6	ABU326003 Protein e
12	1192.5	39.8	576	6	ABM66171 Propionib
13	1191.5	39.7	576	6	ABU45147 Propionib
14	1191.5	39.7	576	6	ABM41666 Propionib
15	1187	39.6	543	6	ABU33806 Protein e
16	1175	39.2	564	6	ABU26242 Protein e
17	1169.5	39.0	581	4	AAG92515 C glutam
18	1169.5	39.0	584	5	AAU97041 Coryneb
19	1169.5	39.0	584	5	ABR79439 Coryneb
20	1169.5	39.0	584	7	ADD13403 C. Glutam
21	1169.5	39.0	584	8	ADJ70036 C glutam
22	1160	38.7	552	4	AAU79569 Coryneb
23	1138.5	38.0	659	5	ABU49819 Listeria
24	1138.5	38.0	659	5	ABU33046 Protein e
25	1098.5	36.6	644	6	ABU18314 Protein e

26	1078.5	35.9	648	4	AAG81983 S. epider
27	1078.5	35.9	662	6	ABU42645 Protein e
28	1078.5	35.9	668	5	ABP38585 Staphyloc
29	1060.5	35.4	662	6	ABU16562 Protein e
30	1060.5	35.4	662	6	ABM73311 Staphyloc
31	1059.5	35.3	649	4	AAU33837 Staphyloc
32	1059.5	35.3	662	4	AAU36750 Staphyloc
33	1056.5	35.2	642	6	ABU23408 Protein e
34	1046.5	34.9	648	6	ABU43487 Protein e
35	1044	34.8	642	6	ABU16881 Protein e
36	1044	34.8	664	6	ADA35769 Aginetoba
37	1040	34.7	660	6	ABU41038 Protein e
38	1040	34.7	672	7	ADF05691 Bacterial
39	1023.5	34.1	646	6	ABU22511 Protein e
40	1020.5	34.0	664	6	ABM67056 Phototrab
41	1019.5	34.0	663	6	ABU50228 Protein e
42	1010.5	33.7	663	4	AAU34477 E. coli c
43	1010.5	33.7	663	4	AAG98862 E. coli g
44	1010.5	33.7	663	6	ABU31973 Protein e
45	1010.5	33.7	663	6	ABU28528 Protein e

ALIGNMENTS

RESULT 1
AAY97750
ID AAY97750 standard; protein; 557 AA.
XX
AC AAY97750;
XX
DT 06-AUG-2001 (first entry)
XX
DE G. oxydans cytochrome C oxidase (COI) protein sequence.

Cytochrome C oxidase; COI; COII; COIII; COI complex; D-sorbitol;
oxidative fermentation; electron transfer; respiratory chain; L-sorbose;
2-Keto-L-gluconic acid; 2KGA production; aldehyde production;
carboxylic acid production; ketone production.
XX Gluconobacter oxydans.
XX EP1103603-A2.
XX 30-MAY-2001.
XX 14-NOV-2000; 2000EP-00124785.
XX 17-NOV-1999; 99EP-00122842.
XX (HOFF) HOFFMANN LA ROCHE & CO AG F.
XX Asakura A, Hoshino T, Shinjoh M;
XX WPI; 2001-357953/38.
XX N-PSDB; AAA91489.
XX New cytochrome c oxidase complex having cytochrome c oxidase activity
from Gluconobacter oxydans DSM 4025, useful in mediating electron
transfer in respiratory chain or producing 2-keto-L-gluconic acid from L-
sorbose or D-sorbitol.
XX Claim 5; Page 16-20; 42pp; English.

This sequence is the Gluconobacter oxydans cytochrome C oxidase (COI) of
the invention. The COI complex is useful in improving oxidative
fermentation and is an essential component mediating electron transfer in
the respiratory chain. The recombinant microorganism and the cytochrome c
oxidase may be used in the genetic preparation of a recombinant COI
complex and in the production of 2-Keto-L-gluconic acid (2KGA) from L-
sorbose or D-sorbitol in a culture medium. The COI is also useful as a
terminal oxidase, oxidising cytochrome c, an electron acceptor from an
enzyme belonging to dehydrogenase for the production of aldehydes,

CC carboxylic acids and ketones from alcohols and aldehydes, especially the
 CC production of 2KGA from L-sorbose or D-sorbitol
 XX
 SQ Sequence 557 AA;

Query Match 100.0%; Score 3000; DB 4; Length 557;
 Best Local Similarity 100.0%; Pred. No. 7.7e-292;
 Matches 557; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MADAAIHGHDEHHEKQGFTHFWSTNKHDIQLLYLVAAGVGVFISVLFTVYMELEMDPG 60
 DB 1 MADAAIHGHDEHHEKQGFTHFWSTNKHDIQLLYLVAAGVGVFISVLFTVYMELEMDPG 60
 QY 61 VQVWCLGARIADASOTCTANGHLNNVMTYHGLMFFVGVIPALFGGFGNYLPLQIG 120
 DB 61 VQVWCLGARIADASOTCTANGHLNNVMTYHGLMFFVGVIPALFGGFGNYLPLQIG 120
 QY 121 APDMAPFRMNNLSFWLFIAGTAMGASLAPAGDGGQLGSGVGVLYPPLSTREAGYSMDL 180
 DB 121 APDMAPFRMNNLSFWLFIAGTAMGASLAPAGDGGQLGSGVGVLYPPLSTREAGYSMDL 180
 QY 181 AIPAVHLSGASSIMGAINMTTFLNMEAPGWLHKVPLFSWSIFITAWLILALPLAGA 240
 DB 181 AIPAVHLSGASSIMGAINMTTFLNMEAPGWLHKVPLFSWSIFITAWLILALPLAGA 240
 QY 241 ITMLLTDNRNFTFPNPAAGGDPILYQHILWFFGHPEVYIIILPGFGIISHVVSTESKPP 300
 DB 241 ITMLLTDNRNFTFPNPAAGGDPILYQHILWFFGHPEVYIIILPGFGIISHVVSTESKPP 300
 QY 301 VFGVLPVYAWAIVGLGVVNAHMYTVGMSLTQOSYFMLATMVAVPTGKIFSWIAT 360
 DB 301 VFGVLPVYAWAIVGLGVVNAHMYTVGMSLTQOSYFMLATMVAVPTGKIFSWIAT 360
 QY 361 MWGGSVEFKSPMLAWAFGMFLFTVGGVTGIVLAQAGLDRAVHDTYVVAHFVYVMSLGA 420
 DB 361 MWGGSVEFKSPMLAWAFGMFLFTVGGVTGIVLAQAGLDRAVHDTYVVAHFVYVMSLGA 420
 QY 421 FAIPAGIYFYMPKFSGRAPFENAAKLHFWTFFIGANVTFFPQHFLGRQGMRRYIDYPA 480
 DB 421 FAIPAGIYFYMPKFSGRAPFENAAKLHFWTFFIGANVTFFPQHFLGRQGMRRYIDYPA 480
 QY 481 FALNWKVSSYGAFLAPASFLFFVIVFYTLVAGRRTRNPNWGEFADTLEWTLPSPPPAH 540
 DB 481 FALNWKVSSYGAFLAPASFLFFVIVFYTLVAGRRTRNPNWGEFADTLEWTLPSPPPAH 540
 QY 541 TPETLPKRSWDKHPSH 557
 DB 541 TPETLPKRSWDKHPSH 557

RESULT 2

ADJ79757
 ID ADJ79757 standard; protein; 514 AA.

XX AC
 XX AC
 XX ADJ79757;
 DT 06-MAY-2004 (first entry)
 XX
 XX Subunit I of bovine cytochrome c oxidase.
 DE cytochrome c oxidase; slimming drug.
 XX Bos taurus.

XX OS
 XX WO2004015095-A1.
 XX PD 19-FEB-2004.
 XX
 XX 02-JUN-2003; 2003WO-JP006912.
 XX
 XX 08-AUG-2002; 2002JP-00231042.
 XX
 XX (NISC-) JAPAN SCI & TECHNOLOGY AGENCY.

XX Yoshikawa S, Shimada H, Shimokata K;
 XX
 XX WPI; 2004-191743/18.
 DR N-PSDB; ADJ79758.

XX Vertebrate cytochrome c oxidase with a modified subunit 1 for potential
 PT use as a slimming drug.
 XX
 XX Disclosure; SEQ ID NO 1; 29pp; Japanese.

XX The present invention relates to a cytochrome c oxidase with a modified
 CC subunit 1 comprising a vertebrate cytochrome c oxidase whose Asp residue
 CC at position 51 from the N-terminal end has been modified. For use in
 CC researching the effects of reduced energy production efficiency in cells
 CC such as nerve and muscle. Potential use as a slimming drug. The present
 CC sequence represents subunit I of bovine cytochrome c oxidase.

XX Sequence 514 AA;

Query Match 51.4%; Score 1543; DB 8; Length 514;

Best Local Similarity 53.2%; Pred. No. 1.1e-145;

Matches 285; Conservative 87; Mismatches 124; Indels 40; Gaps 7;

QY 17 FFRWFSTNKHDIQLLYLVAAGVGVFISVLFTVYMELEMDPGVQVWCLGARIADAS 76
 DB 2 FINRWLFSTNKHDIQLLYLVAAGVGVFISVLFTVYMELEMDPGVQVWCLGARIADAS 76
 QY 77 QFCTANGHLNNVMTYHGLMFFVGVIPALFGGFGNYLPLQIGAPDAPFRMNNLSFWL 136
 DB 46 -TLGDDQIYVNVVTAFAVMIFFVMPVIMIGGFGNWLFLMTIGAPDAPFRMNNLSFWL 104
 QY 137 FIAGTAMGASLAPAGDGGQLGSGVGVLYPPL--STREAGYSMDLAIFAVHLSGASSIM 194
 DB 105 LPPSFLLLASSMV-----EAGAGTGMVYFPPLAGNLAHAGASVDLTIFSLHLAGYSSIL 159
 QY 195 GAINMTTFLNMEAPGWLHKVPLFSWSIFITAWLILALPLAGAITMLLTDNRNFTFP 254
 DB 160 GAINFTIITNMKPPASQYQTLFVWVSMITAVLLLSLFLVLAAGITMLLTDNRNFTFP 219
 QY 255 FNPAGGDPILYQHILWFFGHPEVYIIILPGFGIISHVSTES-KKPVFGVLPVYAWA 313
 DB 220 FDPAGGDPILYQHILWFFGHPEVYIIILPGFGIISHVSTES-KKPVFGVLPVYAWA 279
 QY 314 IGVLFVVAHMYTVGMSLTQOSYFMLATMVAVPTGKIFSWIATMWGGSVEFKSPML 373
 DB 280 IGVLFVVAHMYTVGMSLTQOSYFMLATMVAVPTGKIFSWIATMWGGSVEFKSPML 339
 QY 374 WAFGMFLFTVGGVTGIVLAQAGLDRAVHDTYVVAHFVYVMSLGAIFAFAGIYFMPK 433
 DB 340 WALGFIFFLVGGTGLVLANSSLDIVLHDTYVVAHFVYVMSLGAIFAFAGIYFMPK 399
 QY 434 FSGRAFPENAAKLHFWTFFIGANVTFFPQHFLGRQGMRRYIDYPAFALNWKVSSYGA 493
 DB 400 FSGYTLNDTWAKLHFAIMFVGVNMTFFPQHFLGSGMERNYSYDPDAYTMMWTISSNGSF 459
 QY 494 LAFASFLFFVIVFYTLVAGRR-----TRNPWGEFADTLEWTLPSPPPAH 543
 DB 460 ISLTAVMLMVFI-IWEAFASKREVLTVDLTTN-----LEWLNGCAPPYPTTF 506

RESULT 3

ADE61901
 ID ADE61901 standard; protein; 514 AA.

XX AC
 XX AC ADE61901;
 XX
 DT 29-JAN-2004 (first entry)
 XX
 XX Rat Protein AAB21298, SEQ ID NO 7830.

XX Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;
 KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

XX OS Rattus norvegicus.
XX PN WO2003016475-A2.
XX PD 27-FEB-2003.
XX PF 14-AUG-2002; 2002WO-US025765.
XX PR 14-AUG-2001; 2001US-0312147P.
XX PR 01-NOV-2001; 2001US-0346382P.
XX PR 26-NOV-2001; 2001US-0333347P.
XX PA (GEO) GEN HOSPITAL CORP.
XX PA (FARB) BAYER AG.
XX PI Woolf C, D'urso D, Befort K, Costigan M;
XX DR WPI: 2003-268312/26.
XX DR GENBANK; AAB21296.
XX PT New composition comprising two or more isolated polypeptides, useful for
XX PT preparing a medicament for treating pain in an animal.
XX PS Claim 1; Page; 1017pp; English.
XX CC The invention discloses a composition comprising two or more isolated rat
XX CC or human polynucleotides or a polynucleotide which represents a fragment,
XX CC derivative or allelic variation of the nucleic acid sequence. Also
XX CC claimed are a vector comprising the novel polynucleotide, a host cell
XX CC comprising the vector, a method for identifying a nucleotide sequence
XX CC which is differentially regulated in an animal subjected to pain and a
XX CC kit to perform the method, an array, a method for identifying an agent
XX CC that increases or decreases the expression of the polynucleotide sequence
XX CC that is differentially expressed in neuronal tissue of a first animal
XX CC subjected to pain, a method for identifying a compound which regulates
XX CC the expression of a polynucleotide sequence which is differentially
XX CC expressed in an animal subjected to pain, a method for identifying a
XX CC compound that regulates the activity of one or more of the
XX CC polynucleotides, a method for producing a pharmaceutical composition, a
XX CC method for identifying a compound or small molecule that regulates the
XX CC activity in an animal of one or more of the polypeptides given in the
XX CC specification, a method for identifying a compound useful in treating
XX CC pain and a pharmaceutical composition comprising the one or more
XX CC polypeptides or their antibodies. The polynucleotide or the compound that
XX CC modulates its activity is useful for preparing a medicament for treating
XX CC pain (e.g. spinal segmental nerve injury (SNI)) in an animal (e.g. gene
XX CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
XX CC therapy). The sequence presented is a rat protein (shown in Table 2 of
XX CC the specification) which is differentially expressed during pain. Note:
XX CC The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic form directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 514 AA;
Query Match 51.4%; Score 1542; DB 7; Length 514;
Best Local Similarity 54.2%; Pred. No. 1.4e-145;
Matches 287; Conservative 84; Mismatches 131; Indels 28; Gaps 7;
QY 17 FTRFMSTNKHDKIGLLYLIVAGVGVSISLVTVVNRLELMDPDGVQVYMCLEGARLIADAS 76
DB 2 FVNRMLFSTNKHDKIGLLYLIVAGVGVSISLVTVVNRLELMDPDGVQVYMCLEGARLIADAS 50
QY 77 QTCANGHLNVMVTVHGLNMFVFGIPALFGGFGNLYLMPLOIGAPDMAPPMNNLSFWL 136
DB 51 -----DQYINVIVTAHAFVFMFFVPMVPMVIGFGNVLVPLMIGAPDMAPPMNNLSFWL 104
QY 137 FIAGTANGVASUPAGCGDQGLSGVGWLYPPL--STREAGYSMDLAFVHLSGASSIM 194
DB 105 LPPSFLLLASSMV-----EAGAGTGWTVYPLAGNLAHAGVSDLTIFSLHAGVSIL 159
QY 195 GAINMITTFLNWRAPGWTLLHKVPLFSWSIFITAWLLILLALPLAGAITMLLTDNRFGTTF 254

160 GAINFITTIINKPPAMTQYOTPLFVNSVLIITAVLSSLLSLPVLAAGITMLLTDNRNNTTF 219
255 FNPAGGGDPILOHILWFFGHEVYIIILPGFGIISHVVSTFS-KKPVFGYLPWVYAMVA 313
220 FDPAGGGDPILOHILWFFGHEVYIIILPGFGIISHVVYYSKKKEPGYMGWVWTMS 279
314 IGVLFVWAHMYTVGMSLTQOSYFNLATMVIATVGTGKIFSWIATMGGSGVEFKSPML 373
280 IGFGLFVWAHMYTVGMSLTQOSYFNLATMVIATVGTGKIFSWIATMGGSGVEFKSPML 339
374 WAFGEMLFTVGGVTCIVLAQAGLDRAVHDTYVVAHFHYVMSLGAIFAIFAGIYFMPK 433
340 WALGFILFTVGGTGLIVLSNSSLDIVLHDYVVAHFHYVMSLGAIFAIFAGIYFMPK 399
434 FSGRAFPEWAALKLHFWTFFIGANVTFFPQHFGRQGNPRYIDYPEAFALMNKVSYGAF 493
400 FSGVTLNDTWAKAHFAIMFVGNMTFFPQHFGRQGNPRYIDYPEAFALMNKVSYGAF 459
494 LAPASLFFPIVIFVYTLVAGRRETRNPNWGEFADTLEWTLPSPPPAHTE 543
460 ISLTVL-VNMFVMEAFASKREVLSISYS--STNLEWLHGCPPIHYTE 506

RESULT 4
ABR64214
ID ABR64214 standard; protein; 513 AA.
AC ABR64214;
DT 15-OCT-2003 (first entry)
DE Angiogenesis protein BNO108.
XX Cytostatic; antirheumatic; antiarthritic; antidiabetic; ophthalmological;
XX antipsoriatic; antiarteriosclerotic; cardiant; vasotropic; angiogenesis;
XX gene therapy; vasculature; cancer; rheumatoid arthritis; psoriasis;
XX diabetic retinopathy; cardiovascular disease; atherosclerosis;
XX ischemic limb disease; coronary artery disease.
XX Homo sapiens.
XX WO2003027285-A1.
XX 03-APR-2003.
XX 19-SEP-2002; 2002WO-AU001282.
XX 27-SEP-2001; 2001AU-00007973.
XX 27-SEP-2001; 2001AU-00007974.
XX 11-OCT-2001; 2001AU-00008210.
XX 29-OCT-2001; 2001AU-00008532.
XX 13-NOV-2001; 2001AU-00008838.
XX 28-AUG-2002; 2002AU-00951032.
XX (BION-) BIONOMICS LTD.
XX Gamble JR, Hahn CN, Vadas MA;
XX WPI: 2003-354655/33.
XX N-PSDB; ACP34489.
XX New angiogenic genes and polypeptides, useful for diagnosing,
XX prognosticating or treating an angiogenesis-related disorder, e.g.
XX cancer, rheumatoid arthritis, diabetic retinopathy, psoriasis or
XX cardiovascular diseases.
XX Claim 15; SEQ ID NO 149; 90pp; English.
XX The invention relates to the isolation of novel genes (ACP34446-ACP34559)
XX encoding proteins (ABR64180-ABR64281) involved in the process of
XX angiogenesis. The nucleic acid molecules are useful in identifying and/or
XX obtaining full-length human genes involved in an angiogenic process. The

CC nucleic acid molecule, polypeptides or complexes encoded, cells or
 CC genetically modified non-human animals derived from these are useful for
 CC the screening of candidate pharmaceutical compounds used in treating
 CC angiogenesis-related disorders. They are also useful for diagnosing,
 CC prognosticating or treating an angiogenesis-related disorder, which
 CC involves uncontrolled or enhanced angiogenesis or is a disorder in which
 CC a decreased vasculature is of benefit (e.g. cancer, rheumatoid arthritis,
 CC diabetic retinopathy, psoriasis or cardiovascular diseases such as
 CC atherosclerosis), or involves inappropriately arrested or decreased
 CC angiogenesis or is a disorder in which an expanding vasculature is of
 CC benefit (e.g. ischemic limb disease or coronary artery disease). The
 CC modulator of expression or activity of the polypeptide encoded by the
 CC nucleic acid sequence is useful for manufacturing a medicament for the
 CC treatment of an angiogenesis-related disorder. This sequence corresponds
 CC to one of the novel angiogenic protein
 XX

XX SQ Sequence 513 AA;

Query Match 51.0%; Score 1531; DB 6; Length 513;
 Best Local Similarity 52.4%; Pred. No. 1.8e-144;
 Matches 284; Conservative 81; Mismatches 125; Indels 52; Gaps 6;
 QY 17 FFTWFMSTNKHDKIGLLYLVAAGVGVGIFSVLFTVYMRLEMDPGVQVCMLEGARLIADAS 76
 DB 2 FADRWFSTNKHDKIGLLYLVAAGVGVGIFSVLFTVYMRLEMDPGVQVCMLEGARLIADAS 45
 QY 77 QTCTANGHLNVMVYTHGILMMFFVGIPALFGGFGNVLMPLOIGAPDMAPPRMNNLSFWL 136
 DB 46 -NLGNDDHINVIATAFVIMFFVWPMIIGFGNVLMPLOIGAPDMAPPRMNNLSFWL 104
 QY 137 FIAGTAMGVASLPAGGDLGQSGVGVWLYPPILSTR--EAGYSMDLAIFAHLVSGASSIM 194
 DB 105 LPPSLLLLLASAMV-----EAGAGTGTVPVPLAGNYSHPGASVDLTIFSLHLAGVSSIL 159
 QY 195 GAINMIFTFNMRAPGTLHKVPLFSWISITAWLILLLPVLGAGITMLLTDRNFGTTF 254
 DB 160 GAINFTITINMKPPANTQYOTPLFVWSVLITAVLLLSLPVLAAGITMLLTDRNLTTF 219
 QY 255 FNPAGGDDPILYQHILMFFFGHPEVYIIILPGFGIISHVSVTFS--KKPVFGYLPVYAMVA 313
 DB 220 FDPAGGDDPILYQHILMFFFGHPEVYIIILPGFGIISHVSVTFS--KKPVFGYLPVYAMVA 279
 QY 314 IGVLFVWVAHMYTVMGMSITQGSYFMLATMVAIVPTGKIFSWIATMGGSVFEPKSPML 373
 DB 280 IGVLFVWVAHMYTVMGMSITQGSYFMLATMVAIVPTGKIFSWIATMGGSVFEPKSPML 339
 QY 374 WAFGFMEFLFTVGGVTGIVLAQAGLDRAVDYHDTYVVAHFHYVMSLGAIFAIFAGIYFYMFK 433
 DB 340 WALGFIFLFTVGGTIGIVLANSSLDIVLHDTYVVAHFHYVMSLGAIFAIFAGIYFYMFK 399
 QY 434 FSGRAFPENAAKHLHFTFTFFGANVTFPPQHLGRQGMPPRYIDYDPAFALMKNKVSYGAF 493
 DB 400 FSGYTLDDQYAKHTFTFTFFGANVTFPPQHLGRQGMPPRYIDYDPAFALMKNKVSYGAF 459
 QY 494 LAPAFSLFFVIVFYTLVAGRRTRPNMGEFAD-----TLEWTLFSPPPAHT 541
 DB 460 ISLTAVMLFMFI-----WEAFASKRKLVMLEEPSNVLWLVGCCPPPYHT 504
 QY 542 FE 543
 DB 505 FE 506

RESULT 5

ADL35456
 ID ADL35456 standard; protein; 513 AA.

XX AC ADL35456;

XX DT 03-JUN-2004 (first entry)

XX DE Human cytochrome C oxidase subunit 1 (COX1) protein.

KW angiogenesis; cytostatic; cancer; gene therapy; human;
 KW cytochrome C oxidase subunit 1; COX1; enzyme.

XX Homo sapiens.

XX WO2004019893-A2.

XX PD 11-MAR-2004.

XX 02-SEP-2003; 2003WO-US027523.

XX 30-AUG-2002; 2002US-00231956.

XX (RIGE-) RIGEL PHARM INC.

XX PI Lorens JB, Bogenberger J, Holland S, Xu W;

XX WPI; 2004-239116/22.

XX DR N-PSDB; ADL35455.

XX PT Identifying a compound that modulates angiogenesis for treating e.g.,
 XX cancer comprises contacting the compound with a nucleic acid or
 XX polypeptide and determining the functional effect of the compound upon
 XX the nucleic acid or polypeptide.

XX PS Claim 19; SEQ ID NO 64; 245pp; English.

XX CC The invention relates to a novel method for identifying a compound that
 XX modulates angiogenesis which comprises contacting the compound with a
 XX nucleic acid, or a polypeptide encoded by a nucleic acid, that hybridises
 XX under stringent conditions to a second nucleic acid and determining the
 XX functional effect of the compound upon the nucleic acid or polypeptide.
 XX The method of the invention has cytostatic applications and may be useful
 XX for identifying a compound that modulates angiogenesis for treating
 XX cancer or for use during gene therapy procedures. The current sequence is
 XX that of an angiogenesis-related human protein of the invention.

XX SQ Sequence 513 AA;

Query Match 51.0%; Score 1531; DB 8; Length 513;
 Best Local Similarity 52.4%; Pred. No. 1.8e-144;
 Matches 284; Conservative 81; Mismatches 125; Indels 52; Gaps 6;
 QY 17 FFTWFMSTNKHDKIGLLYLVAAGVGVGIFSVLFTVYMRLEMDPGVQVCMLEGARLIADAS 76
 DB 2 FADRWFSTNKHDKIGLLYLVAAGVGVGIFSVLFTVYMRLEMDPGVQVCMLEGARLIADAS 45
 QY 77 QTCTANGHLNVMVYTHGILMMFFVGIPALFGGFGNVLMPLOIGAPDMAPPRMNNLSFWL 136
 DB 46 -NLGNDDHINVIATAFVIMFFVWPMIIGFGNVLMPLOIGAPDMAPPRMNNLSFWL 104
 QY 137 FIAGTAMGVASLPAGGDLGQSGVGVWLYPPILSTR--EAGYSMDLAIFAHLVSGASSIM 194
 DB 105 LPPSLLLLLASAMV-----EAGAGTGTVPVPLAGNYSHPGASVDLTIFSLHLAGVSSIL 159
 QY 195 GAINMIFTFNMRAPGTLHKVPLFSWISITAWLILLLPVLGAGITMLLTDRNFGTTF 254
 DB 160 GAINFTITINMKPPANTQYOTPLFVWSVLITAVLLLSLPVLAAGITMLLTDRNLTTF 219
 QY 255 FNPAGGDDPILYQHILMFFFGHPEVYIIILPGFGIISHVSVTFS--KKPVFGYLPVYAMVA 313
 DB 220 FDPAGGDDPILYQHILMFFFGHPEVYIIILPGFGIISHVSVTFS--KKPVFGYLPVYAMVA 279
 QY 314 IGVLFVWVAHMYTVMGMSITQGSYFMLATMVAIVPTGKIFSWIATMGGSVFEPKSPML 373
 DB 280 IGVLFVWVAHMYTVMGMSITQGSYFMLATMVAIVPTGKIFSWIATMGGSVFEPKSPML 339
 QY 374 WAFGFMEFLFTVGGVTGIVLAQAGLDRAVDYHDTYVVAHFHYVMSLGAIFAIFAGIYFYMFK 433
 DB 340 WALGFIFLFTVGGTIGIVLANSSLDIVLHDTYVVAHFHYVMSLGAIFAIFAGIYFYMFK 399
 QY 434 FSGRAFPENAAKHLHFTFTFFGANVTFPPQHLGRQGMPPRYIDYDPAFALMKNKVSYGAF 493
 DB 400 FSGYTLDDQYAKHTFTFTFFGANVTFPPQHLGRQGMPPRYIDYDPAFALMKNKVSYGAF 459
 QY 494 LAPAFSLFFVIVFYTLVAGRRTRPNMGEFAD-----TLEWTLFSPPPAHT 541
 DB 460 ISLTAVMLFMFI-----WEAFASKRKLVMLEEPSNVLWLVGCCPPPYHT 504
 QY 542 FE 543
 DB 505 FE 506

Db 400 FSGYTLDTYAKIHFTIMFIGVNLTPFPQHLGSGMPRRYSYDPDAYTTWNILSSVGSF 459
 QY 494 LAFASLEFFIVFVYTLVAGRRTRPNPWGEFAD-----TLEWTLSPPPAHT 541
 Db 460 ISLRAVNLMI FMI-----WEAFASKRKVLVMEEPSMNLWLYGCPPEYHT 504
 QY 542 FE 543
 Db 505 FE 506
 RESULT 6
 ADJ68385
 ID ADJ68385 standard; protein; 513 AA.
 AC ADJ68385;
 XX
 XX 06-MAY-2004 (first entry)
 XX
 XX Human heat mitochondrial protein as a therapeutic target SeqID191.
 XX
 XX mitochondrial; human; screening assay; diabetes mellitus;
 KW Huntington's disease; osteoarthritis;
 KW Leber's hereditary optic neuropathy; LHON;
 KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
 KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;
 KW neuroprotective; nontropic; antidiabetic; anticonvulsant; antiarthritic;
 KW osteopathic; ophthalmological; cytostatic.
 XX
 XX Homo sapiens.
 XX OS
 XX WO2003087768-A2.
 XX
 PD 23-OCT-2003.
 XX
 XX 04-APR-2003; 2003WO-US010870.
 PR 12-APR-2002; 2002US-0372843P.
 PR 17-JUN-2002; 2002US-036987P.
 PR 20-SEP-2002; 2002US-0412418P.
 XX
 XX (MITO-) MITOKOR.
 PA (BUCK-) BUCK INST AGE RES.
 XX
 XX Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;
 PI Warnock DE;
 XX WPI; 2003-845369/78.
 DR
 XX Identifying a mitochondrial target for drug screening assays and for
 PT treating diseases associated with altered mitochondrial function,
 PT comprises detecting a modified polypeptide in a sample and correlating
 PT with the disease.
 XX
 PS Claim 1; SEQ ID NO 191; 180pp; English.
 XX
 XX This invention relates to novel mitochondrial targets that can be used
 CC for therapeutic intervention in treating a disease associated with
 CC altered mitochondrial function. Specifically, it refers to a method for
 CC identifying proteins of the human heart mitochondrial proteome that are
 CC useful for drug screening assays, as well as therapeutic targets. The
 CC present invention describes a method for identifying such proteins that
 CC can be used in the treatment of various diseases associated with altered
 CC mitochondrial function including diabetes mellitus, Huntington's disease,
 CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
 CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
 CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
 CC compositions have neuroprotective, nontropic, antidiabetic,
 CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and
 CC cytostatic activities. This polypeptide sequence is a human heart
 CC mitochondrial protein of the invention.
 XX
 XX Sequence 513 AA;

Query Match 50.9%; Score 1528; DB 7; Length 513;
 Best Local Similarity 52.4%; Pred. No. 3.7e-144;
 Matches 284; Conservative 81; Mismatches 125; Indels 52; Gaps 6;
 QY 17 FFRWFNSTNKHDKIGLLYLVAAGVGVCFISVLFTVVMRLMDPGVOYMCLEGARLADAS 76
 Db 2 FADRWLFSTNKHDKIGLLYLFGNAGVGLGTALSLLRAELGQPG----- 45
 QY 77 QICTANGHLNVMVYVYHGIIMAFVFGIPALFGGFGNYLMPLQIGAPDMAFPRNNLSFWL 136
 Db 46 -NLGNDHIYVIVTAHAFVMIFFVWMPIMIGFGNWLVPVLMIGAPDMAFPRNNNSFWL 104
 QY 137 FIAGTANGVASLFPAGDGGQLSGVGVWLYPPLSTR--EAGYSMDLAIFAVHLSGASSIM 194
 Db 105 LPESLILLASAMV-----EAGAGTGTWYPPPLAGNYSHPGASVDLTFSLHLGAVSSIL 159
 QY 195 GAINMTITFLNMRAPGNTLHKVPLFWSIFITAWLILLALPVLGAGITMLLTDNFGTTF 254
 Db 160 GAINFTITINMKPPANTQYQPLFWNSVLIITAVLLLSLPVLAAGITMLLTDNLTTF 219
 QY 255 FNPAGGGDPLLYOHILWFFGHPEVYIILPGGIIISHVVSTFS-KKPVFGYLPWYAMVA 313
 Db 220 FDPAGGGDPLLYQHLSWFFGHPEVYIILPGFGMISHIVTYYSKKEPFYMGWMMWAMS 279
 QY 314 IGVLFVWVAHHVYTVGMSITQOSYFMLATMTAVPTGKIPSWIATMMGSGVEFKSPML 373
 Db 280 IGLGFLVVAHHNFTVGMVDVTRAITTSMTIITAIPTGVKVSFWLALHSGNNKNSAAVL 339
 QY 374 WAFGFMFLFTVGGVTGIVLAQAGLDRAVDHYVVAHFHYVMSLGAIFAIFAGYFMPK 433
 Db 340 WALGFIFLTVGGLTGIVLANSLSLDIVLHDYVVAHFHYVLSMGAVFAIMGFIHWFPL 399
 QY 434 FSGRAPEWNAKLFHFTFFIGANVTFFPQHFLGQCMPEYIDYDEAFALNKKVSSVGF 493
 Db 400 FSGYTLDTYAKIHFAIMFIGVNLTPFPQHLGSGMPRRYSYDPDAYTTWNILSSVGSF 459
 QY 494 LAFASLEFFIVFVYTLVAGRRTRPNPWGEFAD-----TLEWTLSPPPAHT 541
 Db 460 ISLTAVNLMIFMI-----WEAFASKRKVLVMEEPSMNLWLYGCPPEYHT 504
 QY 542 FE 543
 Db 505 FE 506
 RESULT 7
 ABJ8768
 ID ABJ8768 standard; protein; 530 AA.
 XX
 AC ABJ8768;
 XX
 DT 27-FEB-2003 (first entry)
 XX
 XX Pseudomonas aeruginosa biofilm formation-related protein #32.
 DE
 XX
 KW Biofilm formation modulation; biofilm-associated disease;
 KW cystic fibrosis; AIDS; middle ear infection; acne; periodontal disease;
 KW catheter-associated infection; medical device-associated infection.
 XX
 OS Pseudomonas aeruginosa.
 XX
 XX WO200285295-A2.
 PN
 PD 31-OCT-2002.
 XX
 PF 19-APR-2002; 2002WO-US012532.
 XX
 XX 20-APR-2001; 2001US-0285190P.
 PR 24-OCT-2001; 2001US-0344142P.
 XX
 XX (IOWA) UNIV IOWA RES FOUND.
 PA (HARD) HARVARD COLLEGE.
 PA

XX Whiteley M, Bangera MG, Lory S, Greenberg EP;
 PI
 XX
 DR WPI; 2003-075601/07.
 DR N-PSDB; ABT14590.

XX Identifying compound capable of modulating biofilm formation by
 PT bacteria/bacterial antibiotic resistance, useful for treatment of biofilm
 PT associated disease.

XX Claim 1; Page 116-118; 154pp; English.

XX The invention comprises a method for identifying a compound capable of
 CC modulating biofilm formation by bacteria. The method of the invention is
 CC useful for identifying a compound capable of modulating biofilm formation
 CC by bacteria or modulating bacterial antibiotic resistance. The method of
 CC the invention is also useful for diagnosing and treating a subject
 CC (especially an immunocompromised human) that is afflicted with a biofilm-
 CC associated disease or disorder, such as: cystic fibrosis; AIDS; middle
 CC ear infections; acne; periodontal disease; catheter-associated infections
 CC ; and medical device-associated infections. The present amino acid
 CC sequence represents a protein that is used in the invention

XX Sequence 530 AA;

Query Match 49.0%; Score 1469.5; DB 6; Length 530;
 Best Local Similarity 51.4%; Pred. No. 2.8e-138;
 Matches 279; Conservative 91; Mismatches 146; Indels 27; Gaps 8;

QY 7 HGHDDHE--KQGFTRFMSHNKDKGLLYLVAGVGVISVLFTVYMLELMDPGVQM 64
 DB 11 HAGDHGHPAKGLM-RVLTNTHKIDGTLYLWFSFMMFLGGSMAMVIEAEILFQPLQ-- 67
 QY 65 CLECARLIADASQTCANGHLNMVMTYHILMMFVFGIPALFGFGNLMPLQIGAPDM 124
 DB 68 -----IVEFA-----FFNQMTTHGLINVGAVNFA-FVGLANWIMPLMIGAPDM 111
 QY 125 AFRPMNLSFWLFIAGTANGVASLAPGGDGLGSGVGVLPPLSTREAGYSMDLAIFA 184
 DB 112 ALPRMNFSPWLLPAAFGLLVSTLFMPGG---GNFGMTFYAPLSTTFAPHSVTFEIPA 167
 QY 185 VHLGSSAIMGAINMTITPLNMRAPGWTILHKVPLRFSWIFITAWLILLALPVLAGAIMLT 244
 DB 168 IHLGSSIMGAINVATILNLRAPGWTILMKPLFWTWLITAFLLIAMVPLAGVITNM 227
 QY 245 LTRNFGTTFNPPAGGDPILYQHILWFFGHPEVYIIILPGFGIISHVVVTSKPKVFGY 304
 DB 228 LMDIHFGTSFSAAGGDDPVLFOHVFVFFGHPEVYIMILPAFGAVSAIIPTEARKPLFGY 287
 QY 305 LPMYVAWAIVGLGVFWVAHMYTVGMSLTQOSYFMLATMTAVPTGKIFSWIATMGG 364
 DB 288 TSVVIATASIAFLSFVVAHMFVVGIPVTGELFFMYATMLIAVPTGVKFNWVITMGG 347
 QY 365 SVBFKSPMLWAFGMFLFTVGGTIGVLAQAGLDRAHYDITYVVAHPHYVMSLGAIFAIF 424
 DB 348 SLTFETPLFAVAVILFTIGGSGMLAIPADFOYHDTYFVAHFVYVLPVGAIFGIF 407
 QY 425 AGIYFMPKPSGAPFEWAALHFWTFPGICANVTFFPHQLGROGMPRIYIDPAFALW 484
 DB 408 ASAYYLPKWTGHMYDETGLKLFHMSFTGMNLAFFPMHFVGLAGMPRIPIPNYQFADF 467
 QY 485 NKVSSYGAFIAFASFFFIIVFYVTVLVRRTTRPNMGEFAADTLEWLPSPPAHTPET 544
 DB 468 NMVSSIGAFWEGTQLLFLFI-VIKIRGGKAPAKPW-DGAGLEWSIPSPAPYHTTST 525
 QY 545 LPK 547
 DB 526 PPE 528

RESULT 8
 ABO68563

ID ABO68563 standard; protein; 425 AA.

XX ABO68563;

XX 29-JUL-2004 (first entry)

XX Pseudomonas aeruginosa polypeptide #738.

XX Bacterial infection; Pseudomonas aeruginosa infection; antibacterial.

XX Pseudomonas aeruginosa.

XX US6551795-B1.

XX 22-APR-2003.

XX 18-FEB-1999; 99US-00252991.

XX 18-FEB-1998; 98US-0074788P.

XX 27-JUL-1998; 98US-0094190P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Rubenfield MJ, Nolling J, Deloughery C, Bush D;

XX WPI; 2003-615309/58.

XX N-PSDB; ABD02134.

XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
 PT useful as molecular targets for diagnostics, prophylaxis and treatment of
 PT pathological conditions resulting from bacterial infection.

XX Disclosure; SEQ ID NO 17309; 455pp; English.

XX The invention relates to Pseudomonas aeruginosa polypeptides and the
 CC polynucleotides encoding them. The sequences are useful in diagnosis and
 CC therapy of pathological conditions, as molecular targets for diagnostics,
 CC prophylaxis and treatment of pathological conditions resulting from a
 CC bacterial infection, for evaluating a compound, such as a polypeptide,
 CC for the ability to bind a P. aeruginosa nucleic acid, as components of
 CC effective antibacterial targets, as targets for antibacterial drugs,
 CC including anti-P. aeruginosa drugs, as templates for recombinant
 CC production of P. aeruginosa-derived peptides or polypeptides, as target
 CC components for diagnosis and/or treatment of P. aeruginosa-caused
 CC infection, and in detection of P. aeruginosa sequences or other sequences
 CC of Pseudomonas species using biochip technology. Sequences ABO67826-
 CC ABO84396 represent P. aeruginosa polypeptides of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format from USPTO at
 CC seqdata.uspto.gov/sequence.html

XX Sequence 425 AA;

Query Match 42.6%; Score 1278; DB 7; Length 425;
 Best Local Similarity 55.6%; Pred. No. 3.5e-119;
 Matches 234; Conservative 71; Mismatches 110; Indels 6; Gaps 3;

QY 127 PRMNLSFWLFIAGTANGVASLAPGGDGLGSGVGVLPPLSTREAGYSMDLAIFA 186
 DB 9 PRMNFSFWLLPAAFGLLVSTLFMPGG---GNFGMTFYAPLSTTFAPHSVTFEIPA 64
 QY 187 LSGASSIMGAINMTITPLNMRAPGWTILHKVPLRFSWIFITAWLILLALPVLAGAIMLT 246
 DB 65 LAGTSSIMGAINVATILNLRAPGWTILMKPLFWTWLITAFLLIAMVPLAGVITMMLM 124
 QY 247 DRNFGTTFNPPAGGDPILYQHILWFFGHPEVYIIILPGFGIISHVVVTSKPKVGYLP 306
 DB 125 DIHFGTSFSAAGGDDPVLFOHVFVFFGHPEVYIMILPAFGAVSAIIPTEARKPLFGYTS 184
 QY 307 MYVAWAIVGLGVFWVAHMYTVGMSLTQOSYFMLATMTAVPTGKIFSWIATMGGSV 366
 DB 185 MYTASIAFLSFVVAHMFVVGIPVTGELFFMYATMLIAVPTGVKFNWVITMGGSL 244
 QY 367 BFKSPMLWAFGMFLFTVGGTIGVLAQAGLDRAHYDITYVVAHPHYVMSLGAIFAIFAG 426

Db 245 TFETPLFAFAVFLFTIGGSGMLAIPADFOYHDTYFVAHFHVLVEGAIFGIFAS 304
Qy 427 IYFWMFSGRAPEWAAKLHFWTFITGANVTFPOHFLGRQGMPPRYIDYPEAFALWNK 486
Db 305 AYYLWPKWTHGMYDETGLKLFHFWMSFTOMLAFFPMFVGLAGNPRIPIDYNLQFADFN 364
Qy 487 VSSYGAFLAFASFLFFIVFYITLVAGRRTPRNPWGEADTLEWTLPSPPPAHTPETLP 546
Db 365 VSSIGAFWFGTQLLFLFI-VIKCIRGKGPAPAKPW-DGAEGLWSIFSPAPVHTFTSTPP 422
Qy 547 K 547
Db 423 E 423

RESULT 9
ABU36908
ID ABU36908 standard; protein; 573 AA.
XX
AC ABU36908;
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #22435.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
OS Mycobacterium tuberculosis.
XX
FN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DE WPI; 2003-029926/02.
XX
PS N-PSDB; ACA40778.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.

Claim 25; SEQ ID NO 64832; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX proliferation or the activity of a gene in an operon required for
XX proliferation; (7) identifying a compound that influences the activity of
XX the gene product or that has an activity against a biological pathway; (8)
XX required for proliferation, or that inhibits cellular proliferation; (9)
XX identifying a gene required for cellular proliferation or the biological
XX pathway in which a proliferation-required gene or its gene product lies
XX or a gene on which the test compound that inhibits proliferation of an
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 573 AA;

Query Match 40.8%; Score 1223; DB 6; Length 573;
Best Local Similarity 44.8%; Pred. No. 1.7e-113;
Matches 243; Conservative 87; Mismatches 178; Indels 34; Gaps 9;
Qy 24 STNKHDIILLYVAAGVGVGFIIVLFTVYMELEIMDPGVQVCMLEGARLIADASQTCTANG 83
Db 34 TTDHKMIGIMYCVACISFFIGGLLALLMRTELAAPGLQL-----SN 76
Qy 84 HLWNVMVYHGIILMPEFVGIPALFGFGNYLMPLOIGADPMAPPRMNNLSFWLFIAGTAM 143
Db 77 EQNQLFTHGTHMLLFYATPIVF-GFANVLVLPLOIGADPAFFRLNAPFSLFVFGATI 135
Qy 144 GVASLFAFGDGLGSGVGNWLYPPL--STREAGYSMDLAIFAVHLSGASSINGAINMIT 201
Db 136 GAAGFITPGG-----AADFGWTATPLTDAIHSFGAGGDLWIMGLIVAGLGTILGAVNMIT 191
Qy 202 TFLNMRAPGMLTHKVLPSNSIFITAWLILILALPLVLAGAITMLLTDNRNFTTFNPAGGG 261
Db 192 TVCMRARGMTFRMFIFTNWIMVTSILILIAEPFLTAALFGLAADRHLAGHYDAANG 251
Qy 262 DPILYQHILMFFGHPHYIIILPGGIIISHVSTFKKPVGVLYPMYAMVAIGVLGVV 321
Db 252 -VLLWQHLEFWFGHPHYIIILPFGGIIISHVSTFKKPVGVLYPMYAMVAIGVLGVV 310
Qy 322 WAHMYITVGNLSLTCQSYFEMLATNVIAVPTGKIFPSMTATWGGSVBFKSPMLWAFGMFL 381
Db 311 WAHMFATGAVLL--PFFSFTYLLIAVPTGKIFKNWIGTWKQGLTFETPMLESVGFVMT 368
Qy 382 FTVGVTGIVLAQAGLDRAVDHYVVAHPHYVMSLGAIPAFAGIYFYMKPSGGRAPPE 441
Db 369 FLGLGLTGVLLASPPDLPDHFVTDVYVVAHPHYVLFGTIVFATGAGIYFWFPPKMTGRLLDE 428
Qy 442 WAAKLHFWTFPGICANVTFFPQHFLGRQGMPPRYIDY--PEAFALWNKVSYGAFAPASF 499
Db 429 RLKGLHFWLTFIGFHTTFLVQHMLGDEGMPRRVADYLPDGFQGLNVVSTIGAFILGASM 488
Qy 500 LFFTVIEFVYTLVAGRRTPRNPWGEADTLEWTLPSPPPAHTPETLPK----RSDMDKHP 555
Db 489 PFFVNVFKSWRYGEVTVDDPWG-YGNSLEWATSCPPRRHNFTELPRISERPAPELHY 547
556 SH 557
548 PH 549

RESULT 10
ABU34934
ID ABU34934 standard; protein; 573 AA.

XX
AC ABU34934;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #20461.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.

Claim 25; SEQ ID NO 63927; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *X. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: the sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published/pct/sequences

Sequence 574 AA;

Query Match 40.5%; Score 1215; DB 6; Length 574;

Query Match 40.5%; score 1213; DB 9;
Best Local Similarity 43.7%; Pred. No. 1.1e-112;

Seq. ID	Seq. Name	Length	Score	E-value	Identical	Similarity	Positives	Conservative	Mismatches	Indels	Gaps
1	Seq. 1	100	100.0	0.0	100	100	100	100	0	0	0
2	Seq. 2	100	99.5	0.1	99	99	99	99	1	0	0
3	Seq. 3	100	98.0	0.5	98	98	98	98	2	0	0
4	Seq. 4	100	95.0	1.0	95	95	95	95	5	0	0
5	Seq. 5	100	90.0	5.0	90	90	90	90	10	0	0
6	Seq. 6	100	85.0	10.0	85	85	85	85	15	0	0
7	Seq. 7	100	80.0	20.0	80	80	80	80	20	0	0
8	Seq. 8	100	75.0	50.0	75	75	75	75	25	0	0
9	Seq. 9	100	70.0	100.0	70	70	70	70	30	0	0
10	Seq. 10	100	65.0	200.0	65	65	65	65	35	0	0
11	Seq. 11	100	60.0	500.0	60	60	60	60	40	0	0
12	Seq. 12	100	55.0	1000.0	55	55	55	55	45	0	0
13	Seq. 13	100	50.0	2000.0	50	50	50	50	50	0	0
14	Seq. 14	100	45.0	5000.0	45	45	45	45	55	0	0
15	Seq. 15	100	40.0	10000.0	40	40	40	40	60	0	0
16	Seq. 16	100	35.0	20000.0	35	35	35	35	65	0	0
17	Seq. 17	100	30.0	50000.0	30	30	30	30	70	0	0
18	Seq. 18	100	25.0	100000.0	25	25	25	25	75	0	0
19	Seq. 19	100	20.0	200000.0	20	20	20	20	80	0	0
20	Seq. 20	100	15.0	500000.0	15	15	15	15	85	0	0
21	Seq. 21	100	10.0	1000000.0	10	10	10	10	90	0	0
22	Seq. 22	100	5.0	2000000.0	5	5	5	5	95	0	0
23	Seq. 23	100	1.0	5000000.0	1	1	1	1	99	0	0
24	Seq. 24	100	0.1	10000000.0	0	0	0	0	100	0	0

[illegible]

PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 61730; 1766bp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 543 AA;

Query Match 39.6%; Score 1187; DB 6; Length 543;
 Best Local Similarity 43.8%; Pred. No. 6.6e-110;
 Matches 235; Conservative 86; Mismatches 181; Indels 34; Gaps 9;
 QY 30 IGLLYVAAGVGVFISVLTFTVYMELEMDPGVQVCMLEGARLIADASQCTANGHLNWNM 89
 DB 2 IGIMYTVTCFAFFIGGIMALLMRTSLAAPGLQL-----SNEQFNOL 44
 QY 90 VYHGLIMFVFGIPALPGGNGYMLPLQIGAPDMPAPPMNNLSFWLFIAGTAMGVASLF 149
 DB 45 FTWGHGTMILLYATPVVF-GFANLVLPQIGAPDVAFFRLNAPFSLFLFGGLIAASGFI 103
 QY 150 APGGDGLGSGGVGVLYPLPLS--TREAGYSMDLAIFAVHLSGASSIMGAINMTTFLNMR 207
 DB 104 VPGG----ADFGWTAYTPLSDAVHSPGAGGDLWITGLIVAGLGTILGAVNNITTVCMR 159
 QY 208 APGMTLHKVPLFSWSIFITAWLILLALPVLAGAITMLLTDRNFGTTFFNPAGGDPILYQ 267
 DB 160 APGMTFRMPDIFTNLTIVTILIIAIIAIIAIIAIIAIIAIIAIIAIIAIIAIIAIIAII 218
 QY 268 HILWFFGHPEVYIIILPGFGIIISHVFTSKFVFCGLPMYVAMVAIGLVGVVWAHMY 327
 DB 219 HLFWFFGHPEVYIIALPFGIITTEIPVARKVFGTITLVATLSIALSVAVWAHMY 278
 QY 328 TVGMSLTQSQSYFMLATVIAVPTGIKIFSNIAFMGGSVEFKSPMLWAFGMFLFTVGCV 387
 DB 279 ATGAVLL--PFGSMTYLIAPVTGKIFENWIGTMWKGQLTETEMFLFCVGFLLTFLGGL 336
 QY 388 TGIVLAQAGIDRAYHTYVVAHFYVMSIGATFAIFAGIYFYMPKFSGRAPENAKLH 447
 DB 337 TGVNLASPPLDHFHTYTFVVAHFYVLPFGTIVFATFAGYVFWPMTGRLLDERLGLH 396
 QY 448 FWTFFGANVTFFPQHFLGSGQMPRYIDY--PEAFALWNKVSSYGAFLAFASFLFFVI 505

DB 397 FWLTFIGFHTTFLVQHNLGDLGMPRRYADYLPDSGQFQFYNVASTVGAFILGASMPFFVWN 456
 QY 506 FVYTLVAGRRRETRNPNPWCGEFADTLEWTLPSPPPAHTTETLPK---RSDMDKHPSH 557
 DB 457 VFKSHRYGEVTVDDPWG-YGNSLEWATSCPPPHNTELPRIISRPAPELHYPH 511

Search completed: October 18, 2004, 23:41:10
 Job time : 162 secs

28	124.5	4.2	453	4	US-09-552-991A-32168	Sequence 32168, A
29	123	4.1	119	4	US-09-552-991A-17937	Sequence 17937, A
30	121.5	4.0	336	4	US-09-489-039A-8430	Sequence 8430, Ap
31	121.5	4.0	687	4	US-09-543-681A-5839	Sequence 5839, Ap
32	121.5	4.0	867	4	US-09-107-532A-4393	Sequence 4393, Ap
33	121	4.0	434	4	US-09-328-352-5070	Sequence 5070, Ap
34	120.5	4.0	404	4	US-09-352-991A-18546	Sequence 18546, A
35	120.5	4.0	460	4	US-09-489-039A-13505	Sequence 13505, A
36	120.5	4.0	639	4	US-09-252-991A-17988	Sequence 17988, A
37	120.5	4.0	633	4	US-08-328-352-6163	Sequence 6163, Ap
38	120	4.0	472	4	US-09-489-039A-8121	Sequence 8121, Ap
39	119.5	4.0	533	4	US-09-107-532A-5244	Sequence 5244, Ap
40	117.5	3.9	478	3	US-09-134-001C-5085	Sequence 5085, Ap
41	115.5	3.9	395	4	US-09-489-039A-7739	Sequence 7739, Ap
42	115.5	3.9	419	4	US-09-328-352-4974	Sequence 4974, Ap
43	115.5	3.9	414	4	US-09-107-532A-5141	Sequence 5141, Ap
44	115	3.8	455	4	US-09-252-991A-20384	Sequence 20384, A
45	115	3.8	508	4	US-05-543-661A-6823	Sequence 6823, Ap

ALIGNMENTS

```

; RESULT 1
; US-09-037-889-15
; Sequence 15, Application US/09097889
; Patent No. 6218117
; GENERAL INFORMATION:
; APPLICANT: HerrinStadt, Corrina
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Davis, Robert E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS
; FOR IDENTIFYING
; AGENTS THAT QUANTITATIVELY ALTER DETECTABLE
; MITOCHONDRIAL DNA RATIOS
; TITLE OF INVENTION: EXTRAMITOCHONDRIAL DNA:
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:

```

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1531	51.0	513	3	US-09-097-879-15	Sequence 15, Appl
2	1331	51.0	513	4	US-09-098-079-15	Sequence 15, Appl
3	1278	42.6	425	4	US-09-252-991A-17309	Sequence 17309, A
4	1078.5	35.9	648	4	US-09-710-279-1060	Sequence 1060, Ap
5	1078.5	35.9	668	3	US-09-134-001C-3430	Sequence 3430, Ap
6	1044	34.8	664	4	US-09-328-352-7056	Sequence 7056, Ap
7	1040	34.7	672	4	US-09-543-681A-5976	Sequence 5976, Ap
8	1010.5	33.7	663	4	US-09-711-164-332	Sequence 332, App
9	1010.5	33.7	677	4	US-09-489-039A-13088	Sequence 13088, A
10	974.5	32.5	660	4	US-09-252-991A-22396	Sequence 22396, A
11	234.5	7.8	98	4	US-09-248-796A-16350	Sequence 16350, A
12	186	6.2	472	3	US-09-354-129-2	Sequence 2, Appli
13	186	6.2	472	4	US-09-504-357-2	Sequence 2, Appli
14	163	5.4	161	4	US-09-248-796A-15595	Sequence 15595, A
15	155.5	5.2	474	4	US-09-252-991A-30354	Sequence 30354, A
16	151.5	5.1	423	4	US-09-540-236-3311	Sequence 2311, Ap
17	151.5	5.1	766	4	US-09-248-796A-16626	Sequence 16626, A
18	151	5.0	63	4	US-09-248-796A-14194	Sequence 14194, A
19	150	5.0	623	4	US-09-252-991A-19867	Sequence 19867, A
20	149.5	5.0	538	4	US-09-252-991A-23060	Sequence 23060, A
21	148.5	5.0	61	4	US-09-248-796A-15596	Sequence 15596, A
22	145.5	4.9	501	4	US-09-252-991A-28456	Sequence 28456, A
23	130.5	4.3	507	4	US-09-540-236-3391	Sequence 3391, Ap
24	130.5	4.3	521	4	US-09-540-236-3391	Sequence 2393, Ap
25	128.5	4.3	484	4	US-09-107-532A-3830	Sequence 3830, Ap
26	125	4.2	345	4	US-09-252-991A-23059	Sequence 23059, A
27	125	4.2	417	4	US-09-489-039A-8565	Sequence 8565, A

```

Query Match          51.0%; Score 1531; DB 3; Length 513;
Best Local Similarity 52.4%; Pred. NO. 2.9e-142;
Matches 284; Conservative 81; Mismatches 125; Indels 52; Gaps 6;

17 FTFEWFSTNKHQIGLLYLIVAGWGFGFISVLTVTYMRLMLDPGQYQVMLECARLIADAS 76
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
2 FADPRTTSTYTKVTCITVLYLGAAAGVCTATSLIIRAEVGGPG-----45

```

QY 17 FFTRHFMSTNHKDIGLLYLVAAGVGFISVLFTVMRPELMDDPGVQMCLGARLIADAS 76
| | | | | | | | : : : | |
? FADDPHREPTMVDNITGVLYPCAMAGVGTALSIITPAELGGPG----- 45

QY 77 QTCTANGHLNVMVYHGIILMFFVIGIPALFGGNGVLMPELOIGAPDMAPPRMNNLSFWL 136
 Db 46 -NLLGNDRHIVNIVTAHAFVIMFFVMPIMIGFGNWLVPMTGAPDMAPPRMNNLSFWL 104
 QY 137 FIAGTAMGVASLAPAGGDDGOLGSGVGVLYPPLSTR--EAGYSMDLAIFAHLGSGASSIM 194
 Db 105 LPPSLLLLLASAMV-----EAGAGTGTWVYPPLAGNYSHPGASVDLTIFSLHLAGVSSIL 159
 QY 195 GAINMHTTFLNMRAPGTLHKVPLFSWISITATWLILLLALPVLGAIATMLLTDNRFGTTF 254
 Db 160 GAINFTITLNMKPPATQOTPLFVMSVLITAVLLLSLPLVLAAGITMLLTDNRNLNTTF 219
 QY 255 FNPAGGDDPILYQHILWFFHGFHPEVYIILPGFGIISHVWSTFS--KKPVFGYLPVMYAMVA 313
 Db 220 FDPAGGDDPILYQHILWFFHGFHPEVYIILPGFGMISHIVTYYSKKPEFGYMGVWAMVS 279
 QY 314 IGVLFVWAHMYTVGMSLTQOSYFMLATWVTAFTGKIFSWIATMGGSVFEKSPML 373
 Db 280 IGLGFIIVWAHMYTVGMSLTQOSYFMLATWVTAFTGKIFSWIATMGGSVFEKSPML 373
 QY 374 WAFGFMFLFTVGGTGVIVLAQAGLDRAVHDTYVVAHFHYVMSLGAIFAIFAGIYFMPK 433
 Db 340 WALGFIFLFTVGGTGVIVLAQAGLDRAVHDTYVVAHFHYVMSLGAIFAIFAGIYFMPK 433
 QY 434 FSGRAPPEWAAKLHFWTFFIGANVTFFPQHFLGRQGMPPRYIDYPAFALMKNVSYGAF 493
 Db 400 FSGYTLDDQTYAKIHFTIMFGVNLTPFPQHFLGSGMPRRYSYDPDAYTTWNILSSVGSF 459
 QY 494 LAFASFLFFIVFYVTLVAGRRTNPNMGEFAD-----TLEWTLPSPPAHT 541
 Db 460 ISLTAVMLMIFMI-----WEAFASKKVLVMBEPPSNMNLWLYGCPPPYHT 504
 QY 542 FE 543
 Db 505 FE 506

RESULT 2

US-09-098-079-15
 ; Sequence 15, Application US/09098079
 ; Patent No. 6489095
 ; GENERAL INFORMATION:
 ; APPLICANT: Herrstadt, Corrina
 ; APPLICANT: Ghosh, Soumitra S.
 ; APPLICANT: Cleveland, William
 ; APPLICANT: Fahy, Eoin F.
 ; APPLICANT: Davis, Robert E.
 ; TITLE OF INVENTION: DIAGNOSTIC METHOD BASED ON QUANTIFICATION OF
 ; NUMBER OF SEQUENCES: 26
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED AND BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/098,079
 ; FILING DATE: 15-JUN-1998
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Rosentman Ph.D., Stephen J.
 ; REGISTRATION NUMBER: 43,058
 ; REFERENCE/DOCKET NUMBER: 660088.416
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 15:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 513 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; US-09-098-079-15

Query Match 51.0%; Score 1531; DB 4; Length 513;
 Best Local Similarity 52.4%; Pred. No. 2.9e-142;
 Matches 284; Conservative 81; Mismatches 125; Indels 52; Gaps 6;
 QY 17 FTFWSTNKHDKGLLVLAAGVGVGISVLTFTVYMSLELMDPGVQVMCLGARIADAS 76
 Db 2 FADRWLSTNKHDKGLLVLAAGVGVGISVLTFTVYMSLELMDPGVQVMCLGARIADAS 76
 QY 77 QTCTANGHLNVMVYHGIILMFFVIGIPALFGGNGVLMPELOIGAPDMAPPRMNNLSFWL 136
 Db 46 -NLLGNDRHIVNIVTAHAFVIMFFVMPIMIGFGNWLVPMTGAPDMAPPRMNNLSFWL 104
 QY 137 FIAGTAMGVASLAPAGGDDGOLGSGVGVLYPPLSTR--EAGYSMDLAIFAHLGSGASSIM 194
 Db 105 LPPSLLLLLASAMV-----EAGAGTGTWVYPPLAGNYSHPGASVDLTIFSLHLAGVSSIL 159
 QY 195 GAINMHTTFLNMRAPGTLHKVPLFSWISITATWLILLLALPVLGAIATMLLTDNRFGTTF 254
 Db 160 GAINFTITLNMKPPATQOTPLFVMSVLITAVLLLSLPLVLAAGITMLLTDNRNLNTTF 219
 QY 255 FNPAGGDDPILYQHILWFFHGFHPEVYIILPGFGIISHVWSTFS--KKPVFGYLPVMYAMVA 313
 Db 220 FDPAGGDDPILYQHILWFFHGFHPEVYIILPGFGMISHIVTYYSKKPEFGYMGVWAMVS 279
 QY 314 IGVLFVWAHMYTVGMSLTQOSYFMLATWVTAFTGKIFSWIATMGGSVFEKSPML 373
 Db 280 IGLGFIIVWAHMYTVGMSLTQOSYFMLATWVTAFTGKIFSWIATMGGSVFEKSPML 373
 QY 374 WAFGFMFLFTVGGTGVIVLAQAGLDRAVHDTYVVAHFHYVMSLGAIFAIFAGIYFMPK 433
 Db 340 WALGFIFLFTVGGTGVIVLAQAGLDRAVHDTYVVAHFHYVMSLGAIFAIFAGIYFMPK 433
 QY 434 FSGRAPPEWAAKLHFWTFFIGANVTFFPQHFLGRQGMPPRYIDYPAFALMKNVSYGAF 493
 Db 400 FSGYTLDDQTYAKIHFTIMFGVNLTPFPQHFLGSGMPRRYSYDPDAYTTWNILSSVGSF 459
 QY 494 LAFASFLFFIVFYVTLVAGRRTNPNMGEFAD-----TLEWTLPSPPAHT 541
 Db 460 ISLTAVMLMIFMI-----WEAFASKKVLVMBEPPSNMNLWLYGCPPPYHT 504
 QY 542 FE 543
 Db 505 FE 506

RESULT 3

US-09-252-991A-17309
 ; Sequence 17309, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 17309
 ; LENGTH: 425
 ; TYPE: PRT

ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17309

Query Match 42.6%; Score 1278; DB 4; Length 425;
Best Local Similarity 55.6%; Pred. No. 1.9e-117;
Matches 234; Conservative 71; Mismatches 110; Indels 6; Gaps 3;

QY 127 PRMNNLSFWLPIAGTAMGVASLFAFGGQGLGSGVGVWLYPPLSTREAGYSMDLAIPAVH 186
DB 9 PRMNNFSWLLPAAFGLLVSTLEWEGG---GPNFGWTFYAPLSTTFAPHSVTFPIAH 64
QY 187 LSGASSINGAINMTTFLNMRAPGWLHKVLPFSWIFITAWLLILLALPVLGAIITMLLT 246
DB 65 LAGISSINGAINVIATILNLRAPGWLTKMPLFVMTWLTITAFLLIAYNPVLGAVVTWMLM 124
QY 247 DRNFGTTFNPAAGGDPILYOHILWFFGHPVYIIILPGFGILSHVSTFSKVPVGLP 306
DB 125 DIHFGTSFSAAGGDPVLQHVWFHGFHPEVYIMLPAFGAVSAIIPFARKDLFGYTS 184
QY 307 MYAMVAIGVLGVVVAHHMTVGNLSLTOQSYFMLATMVIAPVTGKIFSWIATMWGGSV 366
DB 185 MYATASIAFLSFVVAHHMFVGPVTPGELFFMYATMLIAVPTGVKVENVTWMEGSL 244
QY 367 EFKSPMLAFGFMELFTVGGVTGIVLAQAGLDRAVHDTYVVAHFHYVMSLGAIFAIPAG 426
DB 245 TFEPMLEFAVAVILFTTGGSGGLMALAPADQVHDTYFVVAHFHYVVLVPGAIFGIFAS 304
QY 427 IVFYPKPSGRAFPWAALKHFWTFFIGANVTFFPQHFLGRQGMPPRYIDYFEAFALMNK 486
DB 305 AYWLPKVTGMYDETCLKLHFWMSFIGWNLAFPMHFPVGLAGMERRIPDYNLQFADFNM 364
QY 487 VSSYGAFIAPASFLPFIYVYTLVAGRRTRPNWGEFADTLEWLPSPPAHTTLP 546
DB 365 VSSIGAFMFGTTLQLFLFI-VIKCIRGGKPAKAPKFW-DGAEGLEWSIFSPAPYHTFTSTPP 422
QY 547 K 547
DB 423 E 423

RESULT 4

US-09-710-279-1060
Sequence 1060, Application US/09710279
Patent No. 6703492
GENERAL INFORMATION:
APPLICANT: KIMMERLY, WILLIAM JOHN
TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
FILE REFERENCE: PU3480US
CURRENT APPLICATION NUMBER: US/09/710,279
PRIOR FILING DATE: 2000-11-09
PRIOR APPLICATION NUMBER: 60/164,258
PRIOR FILING DATE: 1999-11-09
NUMBER OF SEQ ID NOS: 4472
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 1060
LENGTH: 648
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic
OTHER INFORMATION: amino acid sequence
US-09-710-279-1060

Query Match 35.9%; Score 1078.5; DB 4; Length 648;
Best Local Similarity 38.5%; Pred. No. 1.5e-97;
Matches 212; Conservative 102; Mismatches 203; Indels 33; Gaps 8;
QY 18 FTRFMSTNHKDIGLLYLVAAGVGVFISVLTFTYVLMLEMDPGVQVNCLEGARLIADASQ 77
DB 30 YKEWFTSVDHKKIGIMYLISAVLMFVRGGIDALMLRTQLTIPDNKFL----- 76
QY 78 TCTANGHLNVMVYTHGILMFFVIGIPALFGFGNLYLMPLOIGAPDMAPPRMNLSPFLF 137

DB 77 ----EANYHNEVFTTHGVIMIIFMAMFFIF-GLWNVVVIPLQLGARDVAFVPMNVNFWLFL 131
QY 138 IACTANGVASLFAFGGQGLGSGVGVWLYPPLSTR-EAGYSMDLAIFAVHLSGASSIMGA 196
DB 132 FAGMILNLS-FIVGG-----SPAAGWTNYAPLAGEPSGPGVNVYLIAIQISIGSLMTG 186
QY 197 INMITTFLNMRAPGWLHKVLPFSWIFITAWLLILLALPVLGAIITMLLTDRNFGTTFN 256
DB 187 INFVFTILRCKTPTMKFMQMPFVSVTFTTITLIVLAPFVFTVALAMTADRIFGTQFFT 246
QY 257 PAGGDPILYOHILWFFGHPVYIIILPGFGILSHVSTFSKVPVGLPVMYVAVIGV 316
DB 247 VANGGPMNLWANFFVWNGHPEVIVILPAFGMISEIPIFARKRLGHOSMNIWATAGIAP 306
QY 317 LGFVVAHHMTVGNLSLTOQSYFMLATMVIAPVTGKIFSWIATMWGGSVEFKSPMLWAF 376
DB 307 LSLVWVHFFTMGNCALINSFISISTMLIGVPTGVKLENWLLTLYKGRITTESPMLFSL 366
QY 377 GNFELTVGGVTGIVLAQAGLDRAVHDTYVVAHFHYVMSLGAIFAIPAGIYFYMFKFSG 436
DB 367 AFIPNLLGGVTGVMAMASADYQYHNTYFLVAHFHYTLTVGVVFACLAGLIFWYPMGMG 426
QY 437 RAFPWAALKHFWTFFIGANVTFFPQHFLGRQGMPPRYIDYFEAFALM--NKVSSSYCAEL 494
DB 427 YKLNELINKWCFFWFLGNVCFLPQFILGLDGMPPRLTYMPSDCGWLLNFIISTIGAVL 486
QY 495 AFASFLPFIYVYTLVAGRRTRPNWGEFADTLEWLPSP-PPAHTTETLPKRSDDWD- 552
DB 487 MAICFLFLVASIYVSHIKAPREATGDNWDGLRTLEWSTASAIPPKYNAITPDWNYDT 546
QY 553 -----KHPSH 557
DB 547 FVDMKHEGRH 556

RESULT 5

US-09-134-001C-3430
Sequence 3430, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5574
SEQ ID NO 3430
LENGTH: 668
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3430

Query Match 35.9%; Score 1078.5; DB 3; Length 668;
Best Local Similarity 38.5%; Pred. No. 1.6e-97;
Matches 212; Conservative 102; Mismatches 203; Indels 33; Gaps 8;
QY 18 FTRFMSTNHKDIGLLYLVAAGVGVFISVLTFTYVLMLEMDPGVQVNCLEGARLIADASQ 77
DB 50 YKEWFTSVDHKKIGIMYLISAVLMFVRGGIDALMLRTQLTIPDNKFL----- 96
QY 78 TCTANGHLNVMVYTHGILMFFVIGIPALFGFGNLYLMPLOIGAPDMAPPRMNLSPFLF 137
DB 97 ----EANYHNEVFTTHGVIMIIFMAMFFIF-GLWNVVVIPLQLGARDVAFVPMNVNFWLFL 151
QY 138 IACTANGVASLFAFGGQGLGSGVGVWLYPPLSTR-EAGYSMDLAIFAVHLSGASSIMGA 196
DB 152 FAGMILNLS-FIVGG-----SPAAGWTNYAPLAGEPSGPGVNVYLIAIQISIGSLMTG 206

Db 457 GFTLNKXGIRA-FWFWIGFFVAFMPYILGFMGMRRLSQINPE-FHPMLVAAGGA 514
Qy 493 FLAPASLFFIVFYTLVAGR---RETRNPNWGEFADTLEWLPSPPAHTPETLPK-- 547
Db 515 AL-IALGIVCQIQIYVSIRDHLNRLDTGDPWG--GRTLEWSTSSPAPYFAVEPOVK 571
Qy 548 -RSDW 551
Db 572 ARDEW 576

RESULT 8
US-09-711-164-332
; Sequence 332, Application US/09711164
; Patent No. 6589738
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY
; CURRENT APPLICATION NUMBER: US/09/711,164
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-9
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 332
; LENGTH: 663
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-711-164-332

Query Match 33.7%; Score 1010.5; DB 4; Length 663;
Best Local Similarity 37.9%; Pred. No. 7,9e-91;
Matches 212; Conservative 104; Mismatches 193; Indels 51; Gaps 14;
Qy 21 WFMSTNHKXIDIGLLYLVAAGV---GFISVLFTVMRLDLPQVQVCMLEGARLIADSQ 77
Db 48 WLTSDVHKRLGIMYIIVALLRGFADAIM-----MRSQQALASGE 90
Qy 78 TCTANGHLNNVTVYHGLNMFVFCIPALFGGFGNYLMPLOIGAPDMAPPRMNLSPWL 137
Db 91 AGFLPPHYDQFTAHGVIMIFFVAMPFVI-GLMNLVLPQLGARDVAFPELNNLSWFT 149
Qy 138 IAGTAGVASLPAPGGDQGLSGGVGVLYPLPLSTRE--AGYSMDLAIFAVHLSGASSIMG 195
Db 150 VVGVLNVNLSL---GVGEFAQ-TGMLAYPLSGIEYSPGVGVYVWKSLSGIGTTLT 204
Qy 196 AINMTTFLNRPAGMTLHKVLPFNSWIFITAWLILLALPVLGAIATMLLTDRAFGITFF 255
Db 205 GINFFVTILKWRAPGTMFMKVPFTWASLCANVLIIAGFPILTITVALLTLDRLYLGTHFF 264
Qy 256 NPAGGDPITYOHILWFFGHPEVYIIILPGFGIISHVSTFSKPKVFGYLPWYAMVAIG 315
Db 265 TNDMGNMMYINLIWAGHPEVYIIILPVGVFSEIAATFSKELFGYTSILWATVCIT 324
Qy 316 VLGFVVAHMYTVGMSLTQOSYFMLATVIAVPTGIKIFSWIATWVGGSVEFKSPMLWA 375
Db 325 VLSFVWLHHPFTMGAGANVNAFFGTTMIIAIPGVKIFNWLFTMYQGRIVFHSAMLT 384
Qy 376 FGFMEFLFTVGGVTGVLQAGLDRAVHDYVVAHFHYVMSLGATFAIFAGIFYMPKFS 435
Db 385 IGFIIVTFSVGGMTGVLLAVPGADFVHLNLSFLIAHFHNVIIGGVVFGCFAGTYWPKAF 444
Qy 436 GRAFPEWAAKLHFWTFPIGANVTFPPQHFELGRCQMPRR---YIDYPEAFALNKNVSSYGA 492
Db 445 GFKLNETWGRAFWFIIGFFVAFMPLYALGFMGMRRLSQID-PQ-FHTMLMIAASGA 502
Qy 493 FLAPASLFFIVFYTLVAGRRETRP---NPMGEFADTLEWLPSPPAHTPETLP-- 546
Db 503 VL-IALGILCLVQIYVSIRDQNRDLTGDPWG--GRTLEWATSSPPFPYFAVVPVHV 559

Qy 547 -KRSDW-----KHPSH 557
Db 560 ERDAFWEMKEGEAYKPDH 579
RESULT 9
US-09-489-039A-13088
; Sequence 13088, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13088
; LENGTH: 677
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13088

Query Match 33.7%; Score 1010.5; DB 4; Length 677;
Best Local Similarity 37.3%; Pred. No. 8,2e-91;
Matches 210; Conservative 107; Mismatches 195; Indels 51; Gaps 14;
Qy 18 FTFWSTNHKIDIGLLYLVAAGV---GFISVLFTVMLELMDPGVQVCMLEGARLIAD 74
Db 59 WNEWLTSDVHKRLGIMYIIVALLRGFADAIM-----MRSQQVLAS 101
Qy 75 ASOTCTANGHLNNVTVYHGLNMFVFCIPALFGGFGNYLMPLOIGAPDMAPPRMNLSP 134
Db 102 AGEAGLPPHYDQFTAHGVIMIFFVAMPFVI-GLMNLVLPQLGARDVAFPELNNLSF 160
Qy 135 WLIAGTAGVASLPAPGGDQGLSGGVGVLYPLPLSTRE--AGYSMDLAIFAVHLSGASS 192
Db 161 WFTVGVGVILNVNLSL---GVGEFAQ-TGMLAYPLSGIEYSPGVGVYVWKSLSGIGT 215
Qy 193 IMGAINMTTFLNRPAGMTLHKVLPFNSWIFITAWLILLALPVLGAIATMLLTDRAFGT 252
Db 216 TLGIFNFFVTIILKWRAPGTMFMKVPFTWASLCANILIIASFPILTITVALLTDRYLG 275
Qy 253 TFPNAGGDPITYOHILWFFGHPEVYIIILPGFGIISHVSTFSKPKVFGYLPWYAMV 312
Db 276 HFTNDMGNMMYINLIWAGHPEVYIIILPVGVFSEIAATFSKELFGYTSILWATV 335
Qy 313 AIGVLGFVVAHMYTVGMSLTQOSYFMLATVIAVPTGIKIFSWIATWVGGSVEFKSPM 372
Db 336 CIIVLSFIVWLHHPFTMGAGANVNAFFGTTMIIAIPGVKIFNWLFTMYQGRIVNSAM 395
Qy 373 LWAFGMFLFTVGGVTGVLQAGLDRAVHDYVVAHFHYVMSLGAI FAIFAGIFYMP 432
Db 396 MWTIGFIVTFSVGGMTGVLLAVPGADFVHLNLSFLIAHFHNVIIGGVVFGCFAGTYWMP 455
Qy 433 KFSGRAPPEWAAKLHFWTFPIGANVTFPPQHFELGRCQMPRR---YIDYPEAFALNKNVSS 489
Db 456 KAFGFTLNWETWGRAFWFIIGFFVAFMPLYALGFMGMRRLSQID-PQ-FHPMLVIAA 513
Qy 490 YGAPLAFASLFFIVFYTLVAGRRETRP---NPMGEFADTLEWLPSPPAHTPETLP 546
Db 514 CGAALACILCQLIQF-YVSIRDQNRDLTGDPWG--GRTLEWATSSPPFPYFAIVP 570
Qy 547 KRSD-----KHPSH 557
Db 571 QVHERDAFWEMKEGEAYKQPAH 593

RESULT 10
US-09-252-991A-22396
; Sequence 22396, Application US/09252991A

```
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074.788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094.190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22396
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22396

Query Match
32.5%; Score 974.5; DB 4; Length 660;
Best Local Similarity 38.2%; Pred. No. 2.8e-87;
Matches 211; Conservative 97; Mismatches 193; Indels 51; Gaps 15;

QY 18 FTRFWSTNKHDKIGLLYLAAGVYV---GFTSVLFTVYMRLELMDPGVQVCMLEGARLIAD 74
DQ 47 WTEMLTSVDHKKIGVMYIVVALVMLVRGPAD---AIMEGQL-----ALAEAGN---- 92

QY 75 ASOTCTANGHL-----NNVMVYTHGILMMFFVGPALFPGFGNYLMPLOIGAPDMAPPRVN 130
DQ 93 -----HGVLPEPHDQIFTAHGVMIIFMAMP-FMTGLMLAVLPLOIGAPDMAPPRVN 144

QY 131 NLSEFWLFIAGTANGVASLPAFGDGLSGGVGVWLYVPLPSTREAGVS---MDLAIPAVH 186
DQ 145 SLSEFWLVSAMLVNLSL-----GLGEF-ARTGNVAYPLPUS--ELAYSPGVGVYIYIWAQ 197

QY 187 LSGASSIMGAINMITTFLNMRAPGMTLHKVPLFSWISFITAWLILALPVLAGAITMLLT 246
DQ 198 ISGMGLTLLTGINFLVTVKRAPGKMLQMPFTWCTCTEPANILIVASFPILTAAGLLSL 257

QY 247 DRNFGTTFENPAGGDPILYQHLWFFGHPEVYIIILPGFGLIISHVWSTFSKKPVGYLP 306
DQ 258 DRYLDMHFFNEUGGNAMVYINLFWAGHPEVYIILPAGFIPSEVTATFAGKRMFGYKS 317

QY 307 MYVAMVAIGVLGVVWAHMYTVGMSLTQOSYFMLATMVAIVPTGKIFSWIATWGGVS 366
DQ 318 MYVASAATFLGFTVWLHFFHTMGSGDVGNGFFGVATMLISPTGVKLFNMLFTIYKGR 377

QY 367 EFKSPMLWAFGNFLFTVGVGTGIVLAQAQLDRAYHDTYVVAHMYVMSLGAIFAIPAG 426
DQ 378 RFTSTPLWTLGFMVTVTIGGTMGVLLAIPCADFLHNSLFLIAHPHTIIGGAVFGYLAG 437

QY 427 IYFYMFKFSRAPPEWAAKHLFWTFPIGANVTFPFOHFLGROCMRRYIDYPEAFALWKK 486
DQ 438 FAPWPKAAGFTLDEKWKRESFNCWLVGVYMAFPLYLILGFMGTRRLNHYDN--PLMKP 495

QY 487 ---VSSYGAFLAPASFLPIFIVYTLVAGRR---ETRNPNWGEFADTLEWTLPSPPAH 540
DQ 496 YLVYAFPGAVLIFCG-IAOQLQLFVSVRNRKQLADVNGDPWE--CRTLEWATSSPPPFY 552

QY 541 TETLTKRSDWD 552
DQ 553 NFAELPKVQDWD 564

RESULT 11
US-09-248-796A-16350
; Sequence 16350, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
```

```
; CURRENT APPLICATION NUMBER: US/09/248.796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074.725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096.409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 16350
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-16350

Query Match
7.8%; Score 234.5; DB 4; Length 98;
Best Local Similarity 42.3%; Pred. No. 1.7e-15;
Matches 47; Conservative 18; Mismatches 31; Indels 15; Gaps 1;

QY 24 STNKHDKIGLLYLAAGVYVGFISVLFTVYMRLELMDPGVQVCMLEGARLIADASQTCTANG 83
DQ 3 STSHDKIGLLYLYGVMSAVATGMSVILRLSLSGPGMFL-----HGNN 47

QY 84 HLMNVMVYTHGILMMFFVGPALFPGFGNYLMPLOIGAPDMAPPRMNLSE 134
DQ 48 QVFNVLVTGHAIAIMFLFVMPILIGSGFNYFLPIMIGAVDMAFARLNNISF 98

RESULT 12
US-09-354-129-2
; Sequence 2, Application US/09354129
; Patent No. 6136588
; GENERAL INFORMATION:
; APPLICANT: YE, RICK W
; TITLE OF INVENTION: GENES ENCODING DENITRIFICATION REACTIONS
; FILE REFERENCE: CL-1255
; CURRENT APPLICATION NUMBER: US/09/354.129
; CURRENT FILING DATE: 1999-07-15
; EARLIER APPLICATION NUMBER: 60/093.181
; EARLIER FILING DATE: July 17, 1998
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
; TYPE: PRT
; ORGANISM: Pseudomonas sp. strain G-179
US-09-354-129-2

Query Match
6.2%; Score 186; DB 3; Length 472;
Best Local Similarity 21.7%; Pred. No. 8.8e-10;
Matches 98; Conservative 77; Mismatches 173; Indels 104; Gaps 23;

QY 89 MVTYHGLMFFFGIPALFPGFGNYLMPLOIGAPDMAPPRMNLSEFWLFIAGTAMGVASL 148
DQ 51 MIHTNALIVWLLIG-----FMGATVYLLP-EAEETELYSKLAIAQFWIFLAAAIAV-- 103

QY 149 FAPCGDGLSGGVWLYVPLPSTREAGYSMDLAIPAVHLSGASSIMGAINMITTFLNMR 208
DQ 104 -----GYMFKIHGREFEQFIKIGIVV-----CLMFLFNVMTSLKGRK 146

QY 209 PGMTHLKVPLFS-WSIFITAWLILALPVLAGAITMLLTDNRNFTTFFNPAG-GGDPILY 266
DQ 147 TVVT--NILIFGLWGAIP-----FLFSFYVNPANLADKMYW 181

QY 267 QHI--LWFFGHPEVYIILPGF-----GIISHVSTFSKKPVGVKLPVWYAWAI----- 314
DQ 182 WYVVLWLVGEGWELIMASVLAFLMINKLINGIDREVVEKW-----LYVIVGLALFS 230

QY 315 GVLGVVWAHMYTVGM-SLTQOSYFMLATMVAIVPTGKIFSWIATW--GGSVEFKSP 371
DQ 231 GILG---TGHHYVIGAPGVWQWIGLFSLEVPDFTWVITFTVMT-WKAGRKHPNRAA 286

QY 372 MLWAFG-FMFLFTVGGVTGIVLAQAQLDRAYHDTYVVAHFH-----YVWGLGAIFAIF 424
DQ 287 LLWSIGSVMAFFGAGVWGLFHTLSSVNYTHTGTQVTAAGHGLAFLPFGVYVMLNLAIMA-- 344
```


Db 286 INGMWLSGAWHKLRDDPILRFLVSLAFYGMSTFEGPMMAIKTVNSLSHYTDWTIGHVH 345
QY 413 YVMSLGAIFAIFAG-IYFYMPKPSGRAFPENAAKL--HFWTFEIGANVTFFFQHPFLG--R 467
Db 346 -AGALGWANISIGTLYHMIPIKLGREHSHSVGLINAHFWLATIG-TVLYIASMWVNGIT 403
QY 468 QGMPRRYIDYPEAFALWNVKSSYGAFAPASFLFFIVFYTLVA 512
Db 404 QGL-----MWRVNEEDGT-LTYS-----FVEALVA 427

Search completed: October 18, 2004, 23:27:58
Job time : 43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 23:25:00 ; Search time 131 Seconds
(without alignments)
1374.722 Million cell updates/sec

Title: US-09-712-768A-2

Perfect score: 3000

Sequence: 1 MADAAIHGDHHEKQGFTR.....PAHTFETLPKRSWDKHPSH 557

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1360919 seqs, 323318874 residues

Total number of hits satisfying chosen parameters: 1360919

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1531	51.0	513	9	US-09-098-079-15
2	1531	51.0	513	15	Sequence 15, Appl
3	1531	51.0	513	15	Sequence 32, Appl
4	1531	51.0	513	17	US-10-428-487-32
5	1531	51.0	513	17	US-10-231-956A-64
6	1531	51.0	513	17	US-10-770-668-72
7	1531	51.0	513	16	US-10-408-765A-191
8	1531	51.0	513	14	US-10-127-032-117
9	1531	51.0	513	14	US-10-389-647-362
10	1531	51.0	513	14	US-10-156-761-14064
11	1531	51.0	513	14	US-10-282-122A-62858
12	1531	51.0	513	14	US-10-282-122A-64832
13	1531	51.0	513	14	US-10-156-761-13575
14	1531	51.0	513	15	US-10-282-122A-63927
15	1531	51.0	513	15	US-10-282-122A-61730
16	1531	51.0	513	15	US-10-282-122A-54166

15	1169.5	39.0	581	9	US-09-738-626-6269	Sequence 6269, Ap
16	1169.5	39.0	584	9	US-09-945-825-2	Sequence 2, Appli
17	1169.5	39.0	584	15	US-10-380-055-2	Sequence 2, Appli
18	1160	38.7	552	16	US-10-781-014-854	Sequence 654, App
19	1138.5	38.0	659	15	US-10-282-122A-60970	Sequence 60970, A
20	1098.5	36.6	644	15	US-10-282-122A-46238	Sequence 46238, A
21	1078.5	35.9	662	15	US-10-282-122A-70569	Sequence 70569, A
22	1060.5	35.4	662	15	US-10-282-122A-44486	Sequence 44486, A
23	1059.5	35.3	649	9	US-09-815-242-5333	Sequence 5333, Ap
24	1059.5	35.3	662	9	US-09-815-242-12343	Sequence 12343, A
25	1056.5	35.2	642	15	US-10-282-122A-51332	Sequence 51332, A
26	1046.5	34.9	648	15	US-10-282-122A-71411	Sequence 71411, A
27	1044	34.8	642	15	US-10-282-122A-44805	Sequence 44805, A
28	1040	34.7	660	15	US-10-282-122A-68962	Sequence 68962, A
29	1023.5	34.1	646	15	US-10-282-122A-50435	Sequence 50435, A
30	1019.5	34.0	663	15	US-10-282-122A-78152	Sequence 78152, A
31	1010.5	33.7	663	9	US-09-815-242-10070	Sequence 10070, A
32	1010.5	33.7	663	14	US-10-287-274-322	Sequence 322, App
33	1010.5	33.7	663	15	US-10-282-122A-56452	Sequence 56452, A
34	1010.5	33.7	663	15	US-10-282-122A-59897	Sequence 59897, A
35	1004.5	33.5	670	15	US-10-282-122A-69771	Sequence 69771, A
36	1004	33.5	668	15	US-10-282-122A-50550	Sequence 50550, A
37	1003	33.4	668	15	US-10-282-122A-47598	Sequence 47598, A
38	997.5	33.2	663	15	US-10-282-122A-75460	Sequence 75460, A
39	997	33.2	662	15	US-10-282-122A-49239	Sequence 49239, A
40	995.5	33.2	672	15	US-10-282-122A-68178	Sequence 68178, A
41	990.5	33.0	663	9	US-09-815-242-14080	Sequence 14080, A
42	970.5	32.4	658	9	US-09-815-242-11824	Sequence 11824, A
43	970.5	32.4	658	15	US-10-282-122A-66271	Sequence 66271, A
44	970.5	32.4	658	15	US-10-389-647-413	Sequence 413, App
45	957	31.9	652	15	US-10-282-122A-73132	Sequence 73132, A

ALIGNMENTS

RESULT 1

US-09-098-079-15
; Sequence 15, Application US/09098079
; Patent No. US20020064773A1
; GENERAL INFORMATION:
; APPLICANT: Herinstdadt, Corrina
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Cleverger, William
; APPLICANT: Fahy, Eoin F.
; APPLICANT: Davis, Robert E.
; TITLE OF INVENTION: DIAGNOSTIC METHOD BASED ON QUANTIFICATION OF
; TITLE OF INVENTION: EXTRAMITOCHONDRIAL DNA
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,079
; FILING DATE: 15-JUN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rosenman Ph.D., Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 660088.416
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:
 LENGTH: 513 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear

US-09-098-079-15

```

Query Match          51.0%; Score 1531; DB 9; Length 513;
Best Local Similarity 52.4%; Pred. No. 3.1e-126;
Matches 284; Conservative 81; Mismatches 125; Indels 52; Gaps 6;

QY 17 FFRWFMSSTNKHKDIGILLYLVAAGVGFISVLFTVYVMELEMDPGVQVCMLEGARLIADAS 76
Db 2 FADRWLFSTNKHKDIGITLYLLFGAWAGVLTALSLIRAELEGQPG----- 45

QY 77 QTCATANGHLNVMVYHGIILMPFVGIPALFGGFGNLYMPLQIGAPDMAFPNMNLSFWL 136
Db 46 -NLLGNDHIYVIVTAHAFVMIFFVMPIIMIGFGNLYMPLQIGAPDMAFPNMNLSFWL 104

QY 137 FIAGTMGVASLFPAGDGGQGLSGVGVNLYPPLSTR--EAGYSMDLAIFAVHLSGASSTM 194
Db 105 LPPSLLLLLASAMV-----EAGAGTGTWVYPPPLAGNYSHPGASVDLTFSLHLAGVSSIL 159

QY 195 GAINMTITFLNMRAPGMTLHKVPLFSWSIFITAWLILLALPVLAGAITMLLTDNFGTTF 254
Db 160 GAINFITTIINMKPPANTQYQTFLEFVMSVLITAVLLLSLPLVLAAGITMLLTDNLTTF 219

QY 255 FNPAGGDPILYQHILWFHFGHPEVYIILPGFGIISHVSTES--KXPVFGYLPVWYAMVA 313
Db 220 FDPAGGDPILYQHILWFHFGHPEVYIILPGFGIISHVSTES--KXPVFGYLPVWYAMVA 279

QY 314 IGVLFVVAHMYTYGMSLTQOSYFMLATWIAVPTGKIFSWIATMWSGVSEKSPML 373
Db 280 IGVLFVVAHMYTYGMSLTQOSYFMLATWIAVPTGKIFSWIATMWSGVSEKSPML 339

QY 374 WAFGFMFLFTVGGTGVILQAQGLDRAYHDTYVVAHFVYVMSLGAIFAFAFIAGIYFMPK 433
Db 340 WALGFIFLFTVGGTGVILQAQGLDRAYHDTYVVAHFVYVMSLGAIFAFAFIAGIYFMPK 399

QY 434 FSGRAPEWAALKHFTFFIGANVTFPQHFLGRQGMPPRYIDYPEAFALMNKVSYGAF 493
Db 400 FSGYTLDDQTYAKHTFTIMFIGNVLTFFPQHFLGLSGMPRRYSYDPDAYTTNLSVSGSF 459

QY 494 LAFASFLFFIVFYVTLVAGRRETRNPMGEFAD-----TLEWTLPSPPPAHT 541
Db 460 ISITAVMLMIFMI-----WEAFASKRKLVMVEEPSMNLEWLYGCPPPYHT 504

QY 542 FE 543
Db 505 FE 506

```

RESULT 2

```

US-10-428-487-32
; Sequence 32, Application US/10428487
; Publication No. US20040006780A1
; GENERAL INFORMATION:
; APPLICANT: GERBER, HANS-K.
; APPLICANT: GERBER, HANS-PETER
; TITLE OF INVENTION: VEGF-MODULATED GENES AND METHODS EMPLOYING THEM
; FILE REFERENCE: 09800080-0103
; CURRENT APPLICATION NUMBER: US/10/428,487
; CURRENT FILING DATE: 2003-05-02
; PRIOR APPLICATION NUMBER: 09/815,153
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,201
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 32
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Homo sapiens

```

US-10-428-487-32

```

Query Match          51.0%; Score 1531; DB 15; Length 513;
Best Local Similarity 52.4%; Pred. No. 3.1e-126;
Matches 284; Conservative 81; Mismatches 125; Indels 52; Gaps 6;

QY 17 FFRWFMSSTNKHKDIGILLYLVAAGVGFISVLFTVYVMELEMDPGVQVCMLEGARLIADAS 76
Db 2 FADRWLFSTNKHKDIGITLYLLFGAWAGVLTALSLIRAELEGQPG----- 45

QY 77 QTCATANGHLNVMVYHGIILMPFVGIPALFGGFGNLYMPLQIGAPDMAFPNMNLSFWL 136
Db 46 -NLLGNDHIYVIVTAHAFVMIFFVMPIIMIGFGNLYMPLQIGAPDMAFPNMNLSFWL 104

QY 137 FIAGTMGVASLFPAGDGGQGLSGVGVNLYPPLSTR--EAGYSMDLAIFAVHLSGASSTM 194
Db 105 LPPSLLLLLASAMV-----EAGAGTGTWVYPPPLAGNYSHPGASVDLTFSLHLAGVSSIL 159

QY 195 GAINMTITFLNMRAPGMTLHKVPLFSWSIFITAWLILLALPVLAGAITMLLTDNFGTTF 254
Db 160 GAINFITTIINMKPPANTQYQTFLEFVMSVLITAVLLLSLPLVLAAGITMLLTDNLTTF 219

QY 255 FNPAGGDPILYQHILWFHFGHPEVYIILPGFGIISHVSTES--KXPVFGYLPVWYAMVA 313
Db 220 FDPAGGDPILYQHILWFHFGHPEVYIILPGFGIISHVSTES--KXPVFGYLPVWYAMVA 279

QY 314 IGVLFVVAHMYTYGMSLTQOSYFMLATWIAVPTGKIFSWIATMWSGVSEKSPML 373
Db 280 IGVLFVVAHMYTYGMSLTQOSYFMLATWIAVPTGKIFSWIATMWSGVSEKSPML 339

QY 374 WAFGFMFLFTVGGTGVILQAQGLDRAYHDTYVVAHFVYVMSLGAIFAFAFIAGIYFMPK 433
Db 340 WALGFIFLFTVGGTGVILQAQGLDRAYHDTYVVAHFVYVMSLGAIFAFAFIAGIYFMPK 399

QY 434 FSGRAPEWAALKHFTFFIGANVTFPQHFLGRQGMPPRYIDYPEAFALMNKVSYGAF 493
Db 400 FSGYTLDDQTYAKHTFTIMFIGNVLTFFPQHFLGLSGMPRRYSYDPDAYTTNLSVSGSF 459

QY 494 LAFASFLFFIVFYVTLVAGRRETRNPMGEFAD-----TLEWTLPSPPPAHT 541
Db 460 ISITAVMLMIFMI-----WEAFASKRKLVMVEEPSMNLEWLYGCPPPYHT 504

QY 542 FE 543
Db 505 FE 506

```

RESULT 3

```

US-10-231-956A-64
; Sequence 64, Application US/10231956A
; Publication No. US20040053233A1
; GENERAL INFORMATION:
; APPLICANT: Lorens, James B.
; APPLICANT: Xu, Weiduan
; APPLICANT: Bogenberger, Jakob
; APPLICANT: Holland, Sacha
; APPLICANT: Rigol Pharmaceuticals, Incorporated
; TITLE OF INVENTION: Modulators of Angiogenesis
; FILE REFERENCE: 021044-004100US
; CURRENT APPLICATION NUMBER: US/10/231,956A
; CURRENT FILING DATE: 2001-08-30
; NUMBER OF SEQ ID NOS: 522
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 64
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Homo sapiens

```

```

Query Match          51.0%; Score 1531; DB 15; Length 513;
Best Local Similarity 52.4%; Pred. No. 3.1e-126;
Matches 284; Conservative 81; Mismatches 125; Indels 52; Gaps 6;

```

```
QY 17 FTRWFMSTNNHDKDIGLLYLVAAGVVGFIISVLFTVYMRLELMDPGVQVQMCLEGARLIADAS 76
Db 2 FADRWLFSTNNHDKDIGLLYLVAAGVVGFIISVLFTVYMRLELMDPGVQVQMCLEGARLIADAS 45
QY 77 QTCFANGHLNVMVTVYHGIILMMFFVGIIPALFGGNGVLMPLQIGAPDMAPFRMNNSFWL 136
Db 46 -NLLGNDHIYINVTAFHAFVWIFPMWPIIMIGFGNVLVPLMIGAPDMAPFRMNNSFWL 104
QY 137 FIAGTAGVVASLFAFGDGGQGLSGVGVVLYPPLSTR--EAGYSMDLAIFAVHLSGASSIM 194
Db 105 LPPSLLLLLASAMV-----EAGAGTGTVPYPLAGNYSHPGASVDLTIPLSLHLAGVSSIL 159
QY 195 GAINMITTFLNMRAPGMLTKHVPFWSIFITAWLILALPVLGAIITMLLTDNRNFGTTF 254
Db 160 GAINFITTIINKKPAMTQYQTPFLFVMSVLITAVLLLSLPLVLAAGITMLLTDNRNLNTTF 219
QY 255 FNPAGGDDPILYOHILWFFGHPEVYIIILPGFGIISHVVSTFS--KKPVFGVLYPMVYAMVA 313
Db 220 FDPAGGDDPILYOHILWFFGHPEVYIIILPGFGIISHVVSTFS--KKPVFGVLYPMVYAMVA 279
QY 314 IGVLFVVAHMYTVGMSLTQOSVFMATWIAVPTGIKIFSWIATMWGGSVEFKSPML 373
Db 280 IGVLFVVAHMYTVGMSLTQOSVFMATWIAVPTGIKIFSWIATMWGGSVEFKSPML 339
QY 374 WAFGMFLFTVGGTIGVLAQGLDRAYHDTYVVAHPHYVMSLGAIFAIFAGIYFYMFK 433
Db 340 WALGFILFTVGGTIGVLAQGLDRAYHDTYVVAHPHYVMSLGAIFAIFAGIYFYMFK 399
QY 434 FSGRAFPWAALKHFWTFFIGANVTFPQHFLGQGMPPRYIDYPEAFALMNKVSYSYCAF 493
Db 400 FSGYTLDTQYAKIHFTIMEFVGNLTFPQHFLGQGMPPRYIDYPEAFALMNKVSYSYCAF 459
QY 494 LAFASFLFFIVFVYTLVAGRRTRPNWGEFAD-----TLEWTLSPSPPAHT 541
Db 460 ISLTAVMLMIFMI-----WEAFASKRKVLMVEEPSMNLEWLYGCCPPPYHT 504
QY 542 FE 543
Db 505 FE 506

RESULT 4
US-10-770-668-72
; Sequence 72, Application US/10770668
; Publication No. US20040191843A1
; GENERAL INFORMATION:
; APPLICANT: Wright, Susan C.
; APPLICANT: Larrick, James W.
; APPLICANT: Nock, Steven R.
; APPLICANT: Wilson, David S.
; TITLE OF INVENTION: Cell-Killing Molecules and Methods of Use Thereof
; FILE REFERENCE: ABSALUS-08602
; CURRENT APPLICATION NUMBER: US/10/770,668
; CURRENT FILING DATE: 2004-02-02
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 72
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-770-668-72

Query Match 51.0%; Score 1531; DB 17; Length 513;
Best Local Similarity 52.4%; Pred. No. 3.1e-126;
Matches 284; Conservative 81; Mismatches 125; Indels 52; Gaps 6;

QY 17 FTRWFMSTNNHDKDIGLLYLVAAGVVGFIISVLFTVYMRLELMDPGVQVQMCLEGARLIADAS 76
Db 2 FADRWLFSTNNHDKDIGLLYLVAAGVVGFIISVLFTVYMRLELMDPGVQVQMCLEGARLIADAS 45
QY 77 QTCFANGHLNVMVTVYHGIILMMFFVGIIPALFGGNGVLMPLQIGAPDMAPFRMNNSFWL 136
Db 46 -NLLGNDHIYINVTAFHAFVWIFPMWPIIMIGFGNVLVPLMIGAPDMAPFRMNNSFWL 104
```

```
QY 137 FIAGTAGVVASLFAFGDGGQGLSGVGVVLYPPLSTR--EAGYSMDLAIFAVHLSGASSIM 194
Db 105 LPPSLLLLLASAMV-----EAGAGTGTVPYPLAGNYSHPGASVDLTIPLSLHLAGVSSIL 159
QY 195 GAINMITTFLNMRAPGMLTKHVPFWSIFITAWLILALPVLGAIITMLLTDNRNFGTTF 254
Db 160 GAINFITTIINKKPAMTQYQTPFLFVMSVLITAVLLLSLPLVLAAGITMLLTDNRNLNTTF 219
QY 255 FNPAGGDDPILYOHILWFFGHPEVYIIILPGFGIISHVVSTFS--KKPVFGVLYPMVYAMVA 313
Db 220 FDPAGGDDPILYOHILWFFGHPEVYIIILPGFGIISHVVSTFS--KKPVFGVLYPMVYAMVA 279
QY 314 IGVLFVVAHMYTVGMSLTQOSVFMATWIAVPTGIKIFSWIATMWGGSVEFKSPML 373
Db 280 IGVLFVVAHMYTVGMSLTQOSVFMATWIAVPTGIKIFSWIATMWGGSVEFKSPML 339
QY 374 WAFGMFLFTVGGTIGVLAQGLDRAYHDTYVVAHPHYVMSLGAIFAIFAGIYFYMFK 433
Db 340 WALGFILFTVGGTIGVLAQGLDRAYHDTYVVAHPHYVMSLGAIFAIFAGIYFYMFK 399
QY 434 FSGRAFPWAALKHFWTFFIGANVTFPQHFLGQGMPPRYIDYPEAFALMNKVSYSYCAF 493
Db 400 FSGYTLDTQYAKIHFTIMEFVGNLTFPQHFLGQGMPPRYIDYPEAFALMNKVSYSYCAF 459
QY 494 LAFASFLFFIVFVYTLVAGRRTRPNWGEFAD-----TLEWTLSPSPPAHT 541
Db 460 ISLTAVMLMIFMI-----WEAFASKRKVLMVEEPSMNLEWLYGCCPPPYHT 504
QY 542 FE 543
Db 505 FE 506

RESULT 5
US-10-408-765A-191
; Sequence 191, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fany, Bojin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Marnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: PastSeq for Windows Version 4.0
; SEQ ID NO 191
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-191

Query Match 50.9%; Score 1528; DB 16; Length 513;
Best Local Similarity 52.4%; Pred. No. 5.7e-126;
Matches 284; Conservative 81; Mismatches 125; Indels 52; Gaps 6;

QY 17 FTRWFMSTNNHDKDIGLLYLVAAGVVGFIISVLFTVYMRLELMDPGVQVQMCLEGARLIADAS 76
Db 2 FADRWLFSTNNHDKDIGLLYLVAAGVVGFIISVLFTVYMRLELMDPGVQVQMCLEGARLIADAS 45
QY 77 QTCFANGHLNVMVTVYHGIILMMFFVGIIPALFGGNGVLMPLQIGAPDMAPFRMNNSFWL 136
Db 46 -NLLGNDHIYINVTAFHAFVWIFPMWPIIMIGFGNVLVPLMIGAPDMAPFRMNNSFWL 104
QY 137 FIAGTAGVVASLFAFGDGGQGLSGVGVVLYPPLSTR--EAGYSMDLAIFAVHLSGASSIM 194
```

Db 105 LPSLLLLLASAMV-----EAGAGTGWTVYPPPLAGNVSHPGASVDLTIFSLHLAGVSSYL 159
 QY 195 GAINMTTFLNMRAPGMLHKVPLFWSIFITAWLILALPVLAGAITMLLTDNRPGTTF 254
 Db 160 GAINFTIINMKPPANTQOTPLFVNSVLITAVLLLSLPLAAGITMLLTDNRNLNITF 219
 QY 255 FNPAGGDPILYQHILWFFGHPEVYIILPGFGIISHVSTFS-KKPVFGYLPVYAMVA 313
 Db 220 FDPAGGDPILYQHLSWFFGHPEVYIILPGFGWISHVYISGKKEPGYGMWYAMWS 279
 QY 314 IGVLFVVAHMYTVGMSLTQOSYFMLATMYTAVPTGKIFSWIATMVGSGVEKPSML 373
 Db 280 IGLGFIWAHMYTVGMDVDTTAYTSATMIITPTGVKFSWLTATLGSNMKWSAAYL 339
 QY 374 WAFGFMFLFTVGGVTGLVLAOAGLDRAHDYTVVVAHFVYVMSLGAIPALFAGIFYMPK 433
 Db 340 WALGFILFTVGGTLGLVLANSSLDVLHDYTVVVAHFVYVMSLGAIPALFAGIFYMPK 399
 QY 434 FSGRAPPEWAAKLFHFWTFFIGANVTFFPQHFLGRQGMPPRYIDYDPAFALMKNVSSYGAF 493
 Db 400 FSGYTLDTQYAKIHFAIMFICVNLTFPQHFLGLSCMPRRYSDYDPAFALMKNVSSYGAF 459
 QY 494 LAFASPLFFTVVYVTLVAGRRTRPNKGEFAD-----TLEWTLPSPPAHT 541
 Db 460 ISLAVNLMIFMI-----WEAFASKRKLVMVEEPSMNLWLYGCCPPPYHT 504
 QY 542 FE 543
 Db 505 FE 506

RESULT 6

US-10-127-032-117
 ; Sequence 117, Application US/10127032
 ; Publication No. US20030113742A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Whiteley, Marvin
 ; APPLICANT: Bangera, M. Gita
 ; APPLICANT: Lory, Stephen
 ; APPLICANT: Greenberg, Everett Peter
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE MODULATION OF
 ; TITLE OF INVENTION: BIOFILM FORMATION
 ; FILE REFERENCE: UIZ-070CP
 ; CURRENT APPLICATION NUMBER: US/10/127,032
 ; CURRENT FILING DATE: 2002-04-19
 ; PRIOR APPLICATION NUMBER: US 60/285,190
 ; PRIOR FILING DATE: 2001-04-20
 ; PRIOR APPLICATION NUMBER: US 60/344,142
 ; PRIOR FILING DATE: 2001-10-24
 ; NUMBER OF SEQ ID NOS: 170
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 117
 ; LENGTH: 530
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-10-127-032-117

Query Match 49.0%; Score 1469.5; DB 14; Length 530;
 Best Local Similarity 51.4%; Pred. No. 8.6e-121;
 Matches 279; Conservative 91; Mismatches 146; Indels 27; Gaps 8;
 QY 7 HGHDDHHE--KQGFTRFWFSTNHKDIGLLYLVAAGVGVFISVLFTVYMRLEIMDPOVQM 64
 Db 11 HAGDHHGHPAKGLM-RWLTTHNHKDIGTLYLWFSFMFLGGSMAMVIRAEILFQPGQLQ-- 67
 QY 65 CUEGARLIADASQCTANGHLMNVMVYHGIILMFFVGIPLALFGGFGNYLMPLOICAPDM 124
 Db 68 -----IVEPA-----FFNQMTTHGLINMVGAVMPA-FVGLANWMLPLMIGAPDM 111
 QY 125 AFPRMNLFWLFIAGTANGVASLPAFGDGGQLGSGVGWVLYPPLSTRAGYSMDLAIFA 184
 Db 112 ALPRMNFSEWLLPAAFGLLVSTLMPGG-----GNFGWTFYAPLSTTFAPHSVTFEFA 167
 QY 185 VHLGSSIMGAINMTTFLNMRAPGMLHKVPLFWSIFITAWLILALPVLAGAITML 244
 Db 168 IHLAGISSIMGAINMTTFLNMRAPGMLHKVPLFWSIFITAWLILALPVLAGAITML 227

QY 185 VHLGSSIMGAINMTTFLNMRAPGMLHKVPLFWSIFITAWLILALPVLAGAITML 244
 Db 168 IHLAGISSIMGAINMTTFLNMRAPGMLHKVPLFWSIFITAWLILALPVLAGAITML 227
 QY 245 LTRNFGTTFNPNAGGDPILYQHILWFFGHPEVYIILPGFGIISHVSTFSKPKVFGY 304
 Db 228 LMDTHFGTSPFSAAGGDPVLFQHVFWFFGHPEVYIMILPAFGAVSAIIPTPARKPLEGY 287
 QY 305 LPMYVAMVATGVLPVYVAHMYTVGNSLTQOSYFMLATMYTAVPTGKIFSWIATMVG 364
 Db 288 TSMYATASAFUSFVVAHMYTVGNSLTQOSYFMLATMYTAVPTGKIFSWIATMVG 347
 QY 365 SVEPKSMLWAFGFMFLFTVGGVTGLVLAOAGLDRAHDYTVVVAHFVYVMSLGAIPALF 424
 Db 348 SLTETPMLFAVAVILFTTGGFSGMLAIPADQFQVHDYTVVVAHFVYVMSLGAIPALF 407
 QY 425 AGTYNPKFSGRAPPEWAAKLFHFWTFFIGANVTFFPQHFLGRQGMPPRYIDYDPAFALM 484
 Db 408 ASAYYWLPKWTHGMYDETGLKLFHMSFIGMNLAFPPHMFVGLAGMPRRIPDYNLOFADF 467
 QY 485 NKVSSYGAFALAFASFLFFIVFVYVTLVAGRRTRPNKGEFADTLEWTLPSPPAHTFET 544
 Db 468 NMVSSIGAFMFGTITQLLFLI-VIKCIRGKAPAPAKW-DCASGLEWSIESPAPYHTFET 525
 QY 545 LPK 547
 Db 526 PPE 528

RESULT 7

US-10-389-647-362
 ; Sequence 362, Application US/10389647
 ; Publication No. US2004003549A1
 ; GENERAL INFORMATION:
 ; APPLICANT: GREENBERG, E. Peter
 ; APPLICANT: SCHUSTER, Martin
 ; APPLICANT: LOSTROH, Candi
 ; TITLE OF INVENTION: QUORUM SENSING SIGNALING IN BACTERIA
 ; FILE REFERENCE: UIZ-038CP
 ; CURRENT APPLICATION NUMBER: US/10/389,647
 ; CURRENT FILING DATE: 2003-03-14
 ; PRIOR APPLICATION NUMBER: 09/653730
 ; PRIOR FILING DATE: 2000-09-01
 ; PRIOR APPLICATION NUMBER: 60/153022
 ; PRIOR FILING DATE: 1999-09-03
 ; NUMBER OF SEQ ID NOS: 710
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 362
 ; LENGTH: 530
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-10-389-647-362

Query Match 49.0%; Score 1469.5; DB 15; Length 530;
 Best Local Similarity 51.4%; Pred. No. 8.6e-121;
 Matches 279; Conservative 91; Mismatches 146; Indels 27; Gaps 8;

QY 7 HGHDDHHE--KQGFTRFWFSTNHKDIGLLYLVAAGVGVFISVLFTVYMRLEIMDPOVQM 64
 Db 11 HAGDHHGHPAKGLM-RWLTTHNHKDIGTLYLWFSFMFLGGSMAMVIRAEILFQPGQLQ-- 67
 QY 65 CUEGARLIADASQCTANGHLMNVMVYHGIILMFFVGIPLALFGGFGNYLMPLOICAPDM 124
 Db 68 -----IVEPA-----FFNQMTTHGLINMVGAVMPA-FVGLANWMLPLMIGAPDM 111
 QY 125 AFPRMNLFWLFIAGTANGVASLPAFGDGGQLGSGVGWVLYPPLSTRAGYSMDLAIFA 184
 Db 112 ALPRMNFSEWLLPAAFGLLVSTLMPGG-----GNFGWTFYAPLSTTFAPHSVTFEFA 167
 QY 185 VHLGSSIMGAINMTTFLNMRAPGMLHKVPLFWSIFITAWLILALPVLAGAITML 244
 Db 168 IHLAGISSIMGAINMTTFLNMRAPGMLHKVPLFWSIFITAWLILALPVLAGAITML 227

Qy	245	LTDNPGTTFNPAGGDPILYOHILHFGHPEVYIIILPGFGIISHVYSTSKKPEVGY	304
Db	228	LMDIHFGTSFFSAAGGDPVLQHFVWFFGFGHPEVYIIMLFAQAVSAIPTTFARKPLFGY	287
Qy	305	LPMVYMAVIAIGVLGFVVAWHMYTVGMSLTQOQSYFMLATVIAVPTGKIFSNIAIWMGG	364
Db	288	TSMWZATASIAPLSFVVVAWHMFVGVIPVTGELFFMYATMLIAVPTGVKVFVNWVTMMWG	347
Qy	365	SVYEPKSPMLWAFGFMFLFTVGGVTGIVIAQAGLDRAVHDYVYVVAHFYVMSLGAIFAJF	424
Db	348	SLTTFITPMLFAVAVILTTIGFSGMLALAPADFQVHDYTFVVAHFHYVLPVGAIFGIF	407
Qy	425	AGIYFYMFKSGRAPEWAAKLHFWTFFIGANVTFFQHFGLRGQMPPRYIDYDEAPALW	484
Db	408	ASAYVWLPKWTGHMYDETLGLKHLWMSFIGNLAFFPFMHFVGLAGMPRIIPDYNLOFADP	467
Qy	485	NKVSSVGAFAPAFSLFIVIFVYTVLAGRRETRPNPWGEPADLTLEWTLSPPPPAHTFTPT	544
Db	468	NVUSIGAFNFGTTLQLFLFI-VIKYICGGKPAKFW-DGAGGLEWSIPSPAYHTFTPT	525
Qy	545	LPK 547	
Db	526	PPE 528	

RESULT 8

```

US-10-156-761-14064
; Sequence 14064, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDE
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-2040
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-2756
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14064
; LENGTH: 563
; TYPE: PR
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14064

```

[illegible]

```

Db      243  NGG -ALLWOHLFWFFGHPVEVILIAFPFGIITBIIPVSRKPIFGYLTILGATAITGLS 301
Qy      319  FVVVAHHNYTVQMSI.TQOSYEMLATMVIAPVTGIKIFSMIATMWGGSVEFKSPMLWAPCF 378
Db      302  VVVVAHHNFAICAVLL--PPFSFNSFLIAPVTGVKPFNWTGMLKGSLSFEIPMLWATCF 359
Qy      379  MFLFTVGGVGTIVIAQAGLDRAYHDTTVVVAHHYHVMYMSLGAIPAIAGIYFYMPKESGA 438
Db      360  LVSYFLFGLTGVIILASPLDPFHVTDYSFVVAHHYVYVFGVFEATGGFYFWMPKPTGKM 419
Qy      439  FPEWAAKLHFWTFTTIGANVTFFPOHFTLGROGMPRRYIDY--PEAFALMKNVYSYGAFIAF 496
Db      420  LDERLGKIHFWTLFVGPHTTTLVQHWLGAEGMPRRYADYLAADCGFTALNTLSTIGAFLLG 479
Qy      497  ASFLFPIFIVYTLVAGSRETRPMPWGEADTLEWTLPSPPPAHTTETLPK-RSD--WD 552
Db      480  MSTLDFLVNNWKTARYGRKVEZDDPWG-FORSLEWITSCPPPHNFVTLFVRVRESPAD 538
Qy      553  -KPSH 557
Db      539  LHPAH 544

```

RESULT 9

```

US-10-282-122A-62858
; Sequence 62858, Application US/10282122A
; Publication No. US20040029129A1
;
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
;
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;
; FILE REFERENCE: EPIFRA 034A
;
; CURRENT APPLICATION NUMBER: US/10/282,122A
;
; CURRENT FILING DATE: 2003-02-20
;
; PRIOR APPLICATION NUMBER: 60/191,078
;
; PRIOR FILING DATE: 2000-03-21
;
; PRIOR APPLICATION NUMBER: 60/206,848
;
; PRIOR FILING DATE: 2000-05-23
;
; PRIOR APPLICATION NUMBER: 60/207,727
;
; PRIOR FILING DATE: 2000-05-26
;
; PRIOR APPLICATION NUMBER: 60/230,335
;
; PRIOR FILING DATE: 2000-09-06
;
; PRIOR APPLICATION NUMBER: 60/230,347
;
; PRIOR FILING DATE: 2000-09-09
;
; PRIOR APPLICATION NUMBER: 60/242,578
;
; PRIOR FILING DATE: 2000-10-23
;
; PRIOR APPLICATION NUMBER: 60/253,625
;
; PRIOR FILING DATE: 2000-11-27
;
; PRIOR APPLICATION NUMBER: 60/257,931
;
; PRIOR FILING DATE: 2000-12-22
;
; PRIOR APPLICATION NUMBER: 60/267,636
;
; PRIOR FILING DATE: 2001-02-09
;
; PRIOR APPLICATION NUMBER: 60/269,308
;
; PRIOR FILING DATE: 2001-02-16
;
; Remaining Prior Application data removed - See File Wrapper or PALM.
;
; NUMBER OF SEQ ID NOS: 78614
;
; SOFTWARE: PatentIn version 3.1
;
; SEQ ID NO 62858
;
; LENGTH: 573
;
; TYPE: PRT
;
; ORGANISM: Mycobacterium bovis
;
US-10-282-122A-62858

```

RESULT 11
US-10-156-761-13575
; Sequence 13575, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:

Db 369 FLGGTGLVLLSPDLDHVDYFVVAHHYVLCFVIFAGIYFWFPKWTGELLDE 428
Qy 442 WAALKHPTFFIGANVTFFQHFELGRQGMRRYIDY--PEAFALMNKVSYGAFIAPASF 499
Db 429 QLKGKHLFWLTFIGFHTTFLVQHVLGDMGMPRRYADYLPDTGFCGLNVVSTIGSFILGASM 488
Qy 500 LFTFIVFVYTLVAGRETRENPMGEADTLEWTLPPSPPAHFTETLPK---RSDWDKHP 555
Db 489 FPFVWNVKSWRYGEVTVDDPMWG-YGNSLEWATSCPPRRHNTFELPRISRPAFELHY 547
Qy 556 SH 557
Db 548 PH 549

RESULT 13

US-10-282-122A-61730
; Sequence 61730, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61730
; LENGTH: 543
; TYPE: PRT
; ORGANISM: Mycobacterium avium

US-10-282-122A-61730

Query Match 39.8%; Score 1187; DB 15; Length 543;
Best Local Similarity 43.8%; Pred. No. 7e-96;
Matches 235; Conservative 86; Mismatches 181; Indels 34; Gaps 9;

Qy 30 IGLYLVAAGVGVFISVLFVYVVRLELMDPGVQVYMCLEGARLIADASQCTCTANGHLWNVM 89
Db 2 IGIYVTVCFAPFFIGLWALLNRTLEAAGLQFL-----SNEQFNQL 44

Qy 90 VTYHILMMPFVGIPALFGGKGNVLMPLQIGAPDMAEPRNNLSFWLFIAGTANGVASLF 149
Db 45 FTMEGTTMLLLYATPVVF--GFANLVLPQIGAPVAFPRUNAFSFWLFLFGGLAASGFI 103
Qy 150 APGGDGLGSGVGVVLPPLS--TREAGYSMDLAIFAVHLSGASSINGAINMITTFINMR 207
Db 104 VPGG---AADFGWTATPLSDAVHSPGAGGLWITGLIVAGLGTILGAVNMITTVCMR 159
Qy 208 APGMTLHKVPLFSWISITAWLILALPVLGAITMLLTDNFTGTFNPNAGGDPILYQ 267
Db 160 APGMTFRMPIFTWNILVTILILIAFPILTAALFGLAADRHLGAHVYDAANG--VLLWQ 218
Qy 268 HILWFFGHPEVYIIILPGFGIISHVSTFSKPVFGYLPVMYAMVAIGVLGVVVAHHMY 327
Db 219 HLFWFFGHPEVYIIALPFGIITEIIPVFARKPVGYTTLIVYATLSAALSVAVWAHMF 278
Qy 328 TVGMSLTQQSYFEMATWIAVPTGIKIFSIAIATWVGSGVEFKSPMLWAFGFMELFTVGVV 387
Db 279 ATGAVLL--PPFSFMTYLIAVPTGIKFNNWIGTMMKGQLTFETPMLFCVGLLTLILGL 336
Qy 388 TGIVLAQAGLDRAVHDYVYVAHFFHYVMSLGAIFAFAGIYFYMFKFSRAFFPEWAAKLH 447
Db 337 TGVMLASPLDHFVTDYFVVAHFFHYVLCFVIFAGIYFWFPKWTGELLDERLGLKH 396
Qy 448 FTTEFGANVTFFPQHFLGRQGMRRYIDY--PEAFALMNKVSYGAFIAPASFIFVI 505
Db 397 FWLTFIGFHTTFLVQHVLGDMGMPRRYADYLPDGGFQPYNVVASTVGAFILGASMPFVWN 456
Qy 506 FVYTLVAGRETRENPMGEADTLEWTLPPSPPAHFTETLPK---RSDWDKHPSH 557
Db 457 VFKSWRYGEVTVDDPMWG-YGNSLEWATSCPPRRHNTFELPRISRPAFELHYPH 511

RESULT 14

US-10-282-122A-54166
; Sequence 54166, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09

Search completed: October 18, 2004, 23:38:25
Job time : 134 secs

RESULT 15
US-09-738-626-6269
; Sequence 6269, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIICHI
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOT
; FILE REFERENCE: 249-125

GenCore version 5.1.6

QM protein - protein search, using sw model

Run on: October 18, 2004, 23:36:01 ; Search time 42 Seconds
(without alignments)

1276.018 Million cell updates/sec

Title: US-09-712-768A-2

perfect score: 3000

sequence: 1 MADAAIHGHDHHEKQGFFTR.....PAHTFETLPKRSWDWDKHPSH 557

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

total number of hits satisfying chosen parameters: 283416

— 100 —

Maximum DB seq length: 20

1. **Introduction**
 2. **Background**
 3. **Methodology**
 4. **Results**
 5. **Conclusion**
 6. **References**
 7. **Appendix**
 8. **Figure 1**
 9. **Figure 2**
 10. **Figure 3**
 11. **Figure 4**
 12. **Figure 5**
 13. **Figure 6**
 14. **Figure 7**
 15. **Figure 8**
 16. **Figure 9**
 17. **Figure 10**
 18. **Figure 11**
 19. **Figure 12**
 20. **Figure 13**
 21. **Figure 14**
 22. **Figure 15**
 23. **Figure 16**
 24. **Figure 17**
 25. **Figure 18**
 26. **Figure 19**
 27. **Figure 20**
 28. **Figure 21**
 29. **Figure 22**
 30. **Figure 23**
 31. **Figure 24**
 32. **Figure 25**
 33. **Figure 26**
 34. **Figure 27**
 35. **Figure 28**
 36. **Figure 29**
 37. **Figure 30**
 38. **Figure 31**
 39. **Figure 32**
 40. **Figure 33**
 41. **Figure 34**
 42. **Figure 35**
 43. **Figure 36**
 44. **Figure 37**
 45. **Figure 38**
 46. **Figure 39**
 47. **Figure 40**
 48. **Figure 41**
 49. **Figure 42**
 50. **Figure 43**
 51. **Figure 44**
 52. **Figure 45**
 53. **Figure 46**
 54. **Figure 47**
 55. **Figure 48**
 56. **Figure 49**
 57. **Figure 50**
 58. **Figure 51**
 59. **Figure 52**
 60. **Figure 53**
 61. **Figure 54**
 62. **Figure 55**
 63. **Figure 56**
 64. **Figure 57**
 65. **Figure 58**
 66. **Figure 59**
 67. **Figure 60**
 68. **Figure 61**
 69. **Figure 62**
 70. **Figure 63**
 71. **Figure 64**
 72. **Figure 65**
 73. **Figure 66**
 74. **Figure 67**
 75. **Figure 68**
 76. **Figure 69**
 77. **Figure 70**
 78. **Figure 71**
 79. **Figure 72**
 80. **Figure 73**
 81. **Figure 74**
 82. **Figure 75**
 83. **Figure 76**
 84. **Figure 77**
 85. **Figure 78**
 86. **Figure 79**
 87. **Figure 80**
 88. **Figure 81**
 89. **Figure 82**
 90. **Figure 83**
 91. **Figure 84**
 92. **Figure 85**
 93. **Figure 86**
 94. **Figure 87**
 95. **Figure 88**
 96. **Figure 89**
 97. **Figure 90**
 98. **Figure 91**
 99. **Figure 92**
 100. **Figure 93**
 101. **Figure 94**
 102. **Figure 95**
 103. **Figure 96**
 104. **Figure 97**
 105. **Figure 98**
 106. **Figure 99**
 107. **Figure 100**
 108. **Figure 101**
 109. **Figure 102**
 110. **Figure 103**
 111. **Figure 104**
 112. **Figure 105**
 113. **Figure 106**
 114. **Figure 107**
 115. **Figure 108**
 116. **Figure 109**
 117. **Figure 110**
 118. **Figure 111**
 119. **Figure 112**
 120. **Figure 113**
 121. **Figure 114**
 122. **Figure 115**
 123. **Figure 116**
 124. **Figure 117**
 125. **Figure 118**
 126. **Figure 119**
 127. **Figure 120**
 128. **Figure 121**
 129. **Figure 122**
 130. **Figure 123**
 131. **Figure 124**
 132. **Figure 125**
 133. **Figure 126**
 134. **Figure 127**
 135. **Figure 128**
 136. **Figure 129**
 137. **Figure 130**
 138. **Figure 131**
 139. **Figure 132**
 140. **Figure 133**
 141. **Figure 134**
 142. **Figure 135**
 143. **Figure 136**
 144. **Figure 137**
 145. **Figure 138**
 146. **Figure 139**
 147. **Figure 140**
 148. **Figure 141**
 149. **Figure 142**
 150. **Figure 143**
 151. **Figure 144**
 152. **Figure 145**
 153. **Figure 146**
 154. **Figure 147**
 155. **Figure 148**
 156. **Figure 149**
 157. **Figure 150**
 158. **Figure 151**
 159. **Figure 152**
 160. **Figure 153**
 161. **Figure 154**
 162. **Figure 155**
 163. **Figure 156**
 164. **Figure 157**
 165. **Figure 158**
 166. **Figure 159**
 167. **Figure 160**
 168. **Figure 161**
 169. **Figure 162**
 170. **Figure 163**
 171. **Figure 164**
 172. **Figure 165**
 173. **Figure 166**
 174. **Figure 167**
 175. **Figure 168**
 176. **Figure 169**
 177. **Figure 170**
 178. **Figure 171**
 179. **Figure 172**
 180. **Figure 173**
 181. **Figure 174**
 182. **Figure 175**
 183. **Figure 176**
 184. **Figure 177**
 185. **Figure 178**
 186. **Figure 179**
 187. **Figure 180**
 188. **Figure 181**
 189. **Figure 182**
 190. **Figure 183**
 191. **Figure 184**
 192. **Figure 185**
 193. **Figure 186**
 194. **Figure 187**
 195. **Figure 188**
 196. **Figure 189**
 197. **Figure 190**
 198. **Figure 191**
 199. **Figure 192**
 200. **Figure 193**
 201. **Figure 194**
 202. **Figure 195**
 203. **Figure 196**
 204. **Figure 197**
 205. **Figure 198**
 206. **Figure 199**
 207. **Figure 200**
 208. **Figure 201**
 209. **Figure 202**
 210. **Figure 203**
 211. **Figure 204**
 212. **Figure 205**
 213. **Figure 206**
 214. **Figure 207**
 215. **Figure 208**
 216. **Figure 209**
 217. **Figure 210</**

Post-processing: Minimum Match 0%
Minimum Match 100%

Maximum Match 100%
Listing first 45 summaries

4
C
D
C
H
D

```
Database : PLK_79:
1. 1011*
```

1: pir1: *

```
3: pir3:*
```

4. Print:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	2538.5	84.6	558	2	S08270	cytochrome-c oxida
2	2453	81.8	565	2	S20534	cytochrome-c oxida
3	2347	78.2	554	2	S03809	cytochrome-c oxida
4	1897.5	63.2	572	2	A92671	cytochrome-c oxida
5	1890.5	63.0	552	2	A3435	cytochrome-c oxida
6	1873.5	62.5	538	2	S36424	cytochrome-c oxida
7	1847.5	61.6	552	2	D87671	cytochrome-c oxida
8	1829	61.0	541	1	ODZJ1	cytochrome-c oxida
9	1773	59.1	523	2	S62706	cytochrome-c oxida
10	1756.5	58.6	522	2	S25956	cytochrome-c oxida
11	1756.5	58.6	532	2	A97769	cytochrome-c oxida
12	1748.5	58.3	524	2	S14138	cytochrome-c oxida
13	1745	58.2	515	2	T11913	cytochrome-c oxida
14	1744.5	58.1	527	2	S65346	cytochrome-c oxida
15	1744	58.1	534	2	D71698	cytochrome-c oxida
16	1738.5	58.0	527	2	S14139	cytochrome-c oxida
17	1730.5	57.7	524	1	ODRZ1	cytochrome-c oxida
18	1730.5	57.7	528	1	ODZM1	cytochrome-c oxida
19	1729.5	57.6	524	2	S16256	cytochrome-c oxida
20	1713.5	57.1	514	2	S30291	cytochrome-c oxida
21	1707	56.9	537	2	T11237	cytochrome-c oxida
22	1707	56.6	531	2	S78190	cytochrome-c oxida
23	1698.5	56.6	531	2	S65087	cytochrome-c oxida
24	1677	55.9	536	2	S65651	cytochrome-c oxida
25	1670	55.7	533	2	AS9931	cytochrome-c oxida
26	1651.5	55.0	527	1	OBXY1	cytochrome-c oxida
27	1643.5	54.8	527	2	S05290	cytochrome-c oxida
28	1642	54.7	531	2	T12406	cytochrome-c oxida
29	1639	54.6	530	2	T14884	cytochrome-c oxida

Query Match 84.6%; Score 2538.5; DB 2; Length 558;
 Best Local Similarity 81.0%; Pred. No. 7.2e-167;
 Matches 452; Conservative 51; Mismatches 54; Indels 1; Gaps 1;

QY 1 MADAAIHGH-DHBEKQGFTRFMSTNKHDIQLLYLVAAGVGVFISVLFTVMRLMLMDP 59
 DB 1 MADAAIHGH-DHBEKQGFTRFMSTNKHDIQLLYLVAAGVGVFISVLFTVMRLMLMDP 60

QY 60 GVQWMCLEGARLIADASQCTANGHLNWNVYTHGILMMFFVGIIPALFGGF 119
 DB 61 GVQWMCLEGARLIADASQCTANGHLNWNVYTHGILMMFFVGIIPALFGGF 120

QY 120 GAPDMAFPRNNLSFVFLIAGTANGVSLFAPGGDGLGSGVGVWLYPPLSTREAGVMD 179
 DB 121 GAPDMAFPRNNLSYVWYVGVSLAPGGDGLGSGVGVWLYPPLSTREAGVMD 180

QY 180 LAIPAVHLSGASSINGAINMITTFLNRPAGMTLHKVPLPSWIFITAMILLALPVL 239
 DB 181 LAIPAVHLSGASSILGAINITITFLNRPAGMTLHKVPLPSWIFITAMILLALPVL 240

QY 240 AITMLLDRNFQTFPPNAGGDPILYQHILWFEGHPEVYIIILPGFGLSHVSTFSK 299
 DB 241 AITMLLDRNFQTFPPNAGGDPILYQHILWFEGHPEVYIIILPGFGLSHVSTFSK 300

QY 300 PVFGYLPVWYMAVGLGVFWVAHMYTVGMSLTQOSYFMLATMVIAVPTGIKIFSWIA 359
 DB 301 PVFGYLPVWYMAVGLGVFWVAHMYTVGMSLTQOSYFMLATMVIAVPTGIKIFSWIA 360

QY 360 TMWGSVERKSPMLWAFGEFLETVGVGTGIVLAQGLDRAHYDTYVVAHMYVMSLGA 419
 DB 361 TMWGSVERKSPMLWAFGEFLETVGVGTGIVLAQGLDRAHYDTYVVAHMYVMSLGA 420

QY 420 IFAIFAGIYFMPKFSGRAPPEWAAKHEWTFPIGANTVFFPQHFGLGRCMPRRIDYPE 479
 DB 421 IFAIFAGIYFMPKFSGRAPPEWAAKHEWTFPIGANTVFFPQHFGLGRCMPRRIDYPE 480

QY 480 AFALNKNVSSYGAFALFASFLFFVIVTLVAGRETRPNWGPADTLEWTLSPPPA 539
 DB 481 AFALNKNVSSYGAFALFASFLFFVIVTLVAGRETRPNWGPADTLEWTLSPPPA 540

QY 540 HTFETLPKRSDDWKHPSH 557
 DB 541 HTFETLPKRSDDWKHPSH 558

RESULT 2

S20534
 cytochrome-c oxidase (EC 1.9.3.1) chain I - Rhodobacter sphaeroides
 C;Species: Rhodobacter sphaeroides
 C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 20-Aug-1999
 C;Accession: S20534; A42378
 R;Shapleigh, J.P.; Gennis, R.B.
 Mol. Microbiol. 6, 635-642, 1992

A;Title: Cloning, sequencing and deletion from the chromosome of the gene encoding subunit cytochrome-c oxidase (EC 1.9.3.1) chain I - Rhodobacter sphaeroides
 A;Reference number: S20534; MUID:92204019; PMID:1313140
 A;Accession: S20534
 A;Molecule type: DNA
 A;Residues: 1-565 <SH2>
 A;Cross-references: EMBL:X62645; NID:G46417; PIDN:CAA44514.1; PID:G46418
 R;Shapleigh, J.P.; Hill, J.J.; Alben, J.O.; Gennis, R.B.
 J. Bacteriol. 174, 2338-2343, 1992
 A;Title: Spectroscopic and genetic evidence for two heme-Cu-containing oxidases in Rhodospirillum rubrum
 A;Reference number: A42378; MUID:92202164; PMID:1313003
 A;Accession: A42378
 A;Status: Preliminary; nucleic acid sequence not shown; not compared with conceptual translation
 A;Molecule type: DNA
 A;Residues: 286-289, 'L', 292-293, 'G', 295-296, 'MI', 299, 'QIVS', 304-305, 'SR', 308-309, 'V', 311, 'AARLN', 'S', 486, 'S', 488, 'S' <SH2>
 C;Genetics: ctad
 A;Gene: ctad
 C;Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
 C;Keywords: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
 F;24-487/Domain: cytochrome-c oxidase chain I homology <COI>
 F;91,408/Binding site: heme a iron (His) (axial ligands) #status predicted

F;25-500/Domain: cytochrome-c oxidase chain I homology <COI>
 F;102,421/Binding site: heme a iron (His) (axial ligands) #status predicted
 F;284,333,334/Binding site: copper (His) #status predicted
 F;284,333,334/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
 F;288/Binding site: oxygen (Tyr) #status predicted
 F;411/Binding site: magnesium (His) (shared with chain II) #status predicted
 F;419/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 81.8%; Score 2453; DB 2; Length 565;
 Best Local Similarity 78.7%; Pred. No. 5.3e-161;
 Matches 446; Conservative 50; Mismatches 59; Indels 12; Gaps 3;

QY 1 MADAAIHGH-DHBEKQGFTRFMSTNKHDIQLLYLVAAGVGVFISVLFTVMRLMLMDP 60
 DB 1 MADAAIHGH-DHBEKQGFTRFMSTNKHDIQLLYLVAAGVGVFISVLFTVMRLMLMDP 59

QY 61 VQWMCLEGARLIADASQCTANGHLNWNVYTHGILMMFFVGIIPALFGGF 110
 DB 60 VQWMCLEGARLIADASQCTANGHLNWNVYTHGILMMFFVGIIPALFGGF 119

QY 111 GNYLMPLOIGAPDMAFPRNNLSFVFLIAGTANGVSLFAPGGDGLGSGVGVWLYPPLS 170
 DB 120 GNYLMPLOIGAPDMAFPRNNLSYVWYVGVSLAPGGDGLGSGVGVWLYPPLS 179

QY 171 TREAGYSMDLAIPAVHLSGASSINGAINMITTFLNRPAGMTLHKVPLPSWIFITAWLI 230
 DB 180 TREAGYSMDLAIPAVHLSGASSILGAINITITFLNRPAGMTLHKVPLPSWIFITAWLI 239

QY 231 LLALPVLAGAITMLLDRNFQTFPPNAGGDPILYQHILWFEGHPEVYIIILPGFGLIIS 290
 DB 240 LLALPVLAGAITMLLDRNFQTFPPNAGGDPILYQHILWFEGHPEVYIIILPGFGLIIS 299

QY 291 HVSTFESKSPVGVLPVWYMAVGLGVFWVAHMYTVGMSLTQOSYFMLATMVIAVPT 350
 DB 300 HVSTFESKSPVGVLPVWYMAVGLGVFWVAHMYTVGMSLTQOSYFMLATMVIAVPT 359

QY 351 GIKIFSWIATMWGSEVFKSPMLWAFGEFLETVGVGTGIVLAQGLDRAHYDTYVVAH 410
 DB 360 GIKIFSWIATMWGSEVFKSPMLWAFGEFLETVGVGTGIVLAQGLDRAHYDTYVVAH 419

QY 411 FHYVMSLGAIFAFAGIYFMPKFSGRAPPEWAAKHEWTFPIGANTVFFPQHFGLGRCQM 470
 DB 420 FHYVMSLGAIFAFAGIYFMPKFSGRAPPEWAAKHEWTFPIGANTVFFPQHFGLGRCQM 479

QY 471 PRRYIDYPEAFALNKNVSSYGAFALFASFLFFVIVTLVAGRETRPNWGPADTLE 530
 DB 480 PRRYIDYPEAFALNKNVSSYGAFALFASFLFFVIVTLVAGRETRPNWGPADTLE 538

QY 531 WTLTSPPTTLPKRSDDWKHPSH 557
 DB 539 WTLTSPPTTLPKRSDDWKHPSH 565

RESULT 3

S03809
 cytochrome-c oxidase (EC 1.9.3.1) chain I - Paracoccus denitrificans
 C;Species: Paracoccus denitrificans
 C;Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 09-Jul-2004
 C;Accession: S03809
 R;Raitio, M.; Jalli, T.; Saraste, M.
 EMBO J. 6, 2825-2833, 1987

A;Title: Isolation and analysis of the genes for cytochrome c oxidase in Paracoccus denitrificans
 A;Reference number: S03803
 A;Accession: S03809
 A;Molecule type: DNA
 A;Residues: 1-554 <RAI>
 A;Cross-references: UNIPROT:P08305; EMBL:X05829; NID:G994806; PIDN:CAA29274.1; PID:G9956;
 C;Genetics:
 A;Gene: COI
 C;Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
 C;Keywords: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
 F;24-487/Domain: cytochrome-c oxidase chain I homology <COI>
 F;91,408/Binding site: heme a iron (His) (axial ligands) #status predicted

F:273,322,323/Binding site: copper (His) #status predicted
F:273-277/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
F:277/Binding site: oxygen (Tyr) #status predicted
F:398/Binding site: magnesium (His) (shared with chain II) #status predicted
F:406/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 78.2%; Score 2347; DB 2; Length 554;
Best Local Similarity 76.1%; Pred. No. 9.8e-154;
Matches 415; Conservative 59; Mismatches 69; Indels 2; Gaps 1;

QY 13 EKQGFTRFWSTNHNKIDIGLLYVAAGVGVFISVLTVMRLMDPGVQVMCLGARLI 72
DB 11 EKRGFFTFWSTNHNKIDIGLLYVFTAGLAGLISVLTVMRLMDPGVQVMCLGARLV 70

QY 73 ADASOTCTANGHLNVMVYTHGILMMPFVGIPALFGGFGNYLMBLQIADPMAFPRMNL 132
DB 71 ADAARECTPNALNVMVYTHGILMMPFVGIPALFGGFGNYLMBLQIADPMAFPRMNL 130

QY 133 SFWLPFIAGTANGVASLFPAGGDDGQSGVGMVLPPLSTREAGYSMDLAIPAVHLSGASS 192
DB 131 SYWLVCVGSIAIASLLSPGSDQPGAGVGNVLPPLSTREAGYAMDIAIFAVHVSATS 190

QY 193 IMGAINMTTFLNMRAPGWLTKVPLFSWSTITAWLILALPVLAGAITMLLTDNRFGT 252
DB 191 ILGAINMTTFLNMRAPGWLTKVPLFAVAFITAMILLSLPVLAGGITMLLMDNRFGT 250

QY 253 TFEHPAGGDDPILYQHILWFHFGHPEVYIILPGFGIISHVYSTTSKPVFGYLPVMYAMV 312
DB 251 QFFDPAGGDDPVLVYQHILWFHFGHPEVYIILPGFGIISHVYSTTSKPVFGYLPVMYAMA 310

QY 313 AIGVLGFWVAAHMYTVGMSLTQSQSYFMLATMVAIVPTGKIFSWIATMCGSVFSPKSPM 372
DB 311 AIAPLGFIWAAHMYTVGMSLTQSQSYFQWATMTAVPTGKIFSWIATMCGSVFSPKSPM 370

QY 373 LWAFGFMELFTVGGVTGIVLAQAGLDRAVHDTYVVAHVHVMVSLGAIFAFAGIYFMP 432
DB 371 LWALA--FLFTVGGVTGVVIAQGLSDRVYHDTYVVAHVHVMVSLGALFAIFAGIYWI 428

QY 433 KPSGRAFFEWAAKLFHFTFFIGANTVFPQFLGRQGMRRYIDYPEAFALNKNVSSYGA 492
DB 429 KMSGRQYEWAGQLHFWMMFGSNLIFFPQHLGRQGMRRYIDYPEAFVYNNISSIGA 488

QY 493 FLAFASLFFIVFYVTLVAGRRTRPNPWEFADTLWNTLPSPPAHFTTLTKRSDWD 552
DB 489 YISPASLFFIGIVFYVTLVAGRRTRPNPWEFADTLWNTLPSPPAHFTTLTKRSDWD 548

QY 553 KHPSH 557
DB 549 RAQAH 553

RESULT 4
AB2671
Cytochrome-c oxidase chain I cox4 [imported] - Agrobacterium tumefaciens (strain C58, Du
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
R:Accession: AB2671
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo, I.
exage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
sner, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AB2671
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-572 <KUR>
A:Cross-references: UNIPROT:Q0UHH4; GB:AE008688; PIDN:AAL41784.1; PID:gl7739138; GSPDB:C
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: cox4

A:Map position: circular chromosome
C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
F:299,348,349/Binding site: copper (His) #status predicted
F:299-303/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
F:303/Binding site: oxygen (Tyr) #status predicted
F:426/Binding site: magnesium (His) (shared with chain II) #status predicted
F:434/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 63.2%; Score 1897.5; DB 2; Length 572;
Best Local Similarity 63.9%; Pred. No. 7e-123;
Matches 350; Conservative 73; Mismatches 112; Indels 13; Gaps 6;

QY 7 HGHDEHEKGGFFTRFWSTNHNKIDIGLLYVAAGVGVFISVLTVMRLMDPGVQVMCL 66
DB 29 HSDHSHKGFARWFLSTNHNKIDIGLLYVFAIMAGIIGGGLSVNRMELQEPGQI--LF 86

QY 67 EG-ARLIADASOTCTANG--HLNVMVYTHGILMMPFVGIPALFGGFGNYLMBLQIADP 123
DB 87 HGLASVYGFEGDAADIGGKHNPNVFTTAHALIMIFFVMNPAMIGGFANWMIPIGAPD 146

QY 124 MAPRMNLSFWLFIAGTANGVASLFPAGGDDGQSGVGMVLPPLSTR-EAGYSMDLAI 182
DB 147 MAPRLNLSFWLFIAGTANGVASLFPAGGDDGQSGVGMVLPPLSTRSGMGPVADLAI 206

QY 183 FAVHLSGASSIMGAINMTTFLNMRAPGWLTKVPLFSWSTITAWLILALPVLAGAIT 242
DB 207 FSLHVAGASSILGAINMTTFLNMRAPGWLTKVPLFSWSTITAWLILALPVLAGAIT 266

QY 243 MLITDRNFGTTFNPPAGGDDPILYQHILWFHFGHPEVYIILPGFGIISHVYSTTSKPV 302
DB 267 MLITDRNFGTTFNPPAGGDDPILYQHILWFHFGHPEVYIILPGFGIISHVYSTTSKPV 336

QY 303 GYLPVMYAMVAIGVLGFWVAAHMYTVGMSLTQSQSYFMLATMVAIVPTGKIFSWIATM 362
DB 327 GYLGMAYAMVAIGVFIWAAHMYTVGMSLTQSQSYFMLATMVAIVPTGKIFSWIATM 386

QY 363 GGSVEFKSPMLAFGNFLFTVGGVTGIVLAQAGLDRAVHDTYVVAHVHVMVSLGAIFA 422
DB 387 GGSULTSTPMVAIGRIFLFTVGGVTGIVLAQAGLDRAVHDTYVVAHVHVMVSLGAVEA 446

QY 423 IFAGIYFMPKPSGRAFFEWAAKLFHFTFFIGANTVFPQFLGRQGMRRYIDYPEAF 482
DB 447 IFAGIYFMPKPSGRAFFEWAAKLFHFTFFIGANTVFPQFLGRQGMRRYIDYPEAF 506

QY 483 LMKVSSYGAFL--AFASLFFIVFYVTLVAGRRTRPNPWEFADTLWNTLPSPPA 539
DB 507 GNNVSSYGSYISAVAVGIFL----GVWEAFAKKRIAGNNPMGEGATTLEMLSSPP 562

QY 540 HTPETLPK 547
DB 563 HQWEQLPR 570

RESULT 5
AC3435
Cytochrome-c oxidase (EC 1.9.3.1) [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C:Accession: AC3435
R:DelVecchio, V.G.; Kapral, R.J.; Redkar, R.J.; Patra, G.; Mujer, C.; Ivanova, I.
; Nazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AC3435
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-552 <KUR>
A:Cross-references: UNIPROT:Q8YFQ5; GB:AE008917; PIDN:AAL52646.1; PID:gl7983469; GSPDB:C
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI1465

A;Map position: I
 C;Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
 C;Keywords: copper; electron transfer; heme; iron; magnesium; membrane-associated complex
 F;279-328/Binding site: copper (His) #status predicted
 F;279-283/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
 F;283/Binding site: oxygen (Tyr) #status predicted
 F;406/Binding site: magnesium (His) (shared with chain II) #status predicted
 F;414/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 63.0%; Score 1890.5; DB 2; Length 552;
 Best Local Similarity 62.8%; Pred. No. 2e-122;
 Matches 350; Conservative 77; Mismatches 113; Indels 17; Gaps 8;

QY 1 MADAAIH---CHDHEKQGFTHFWSTNHNKIDGLLYLVAAGVGVISLVFTVYMLELM 57
 DB 1 MAGTAAHEGAHDHDKPHG-WWRVWYSTNHNKIDGLTYLFAITAGTIGGALSIAEAEIQ 59
 QY 58 DPGVQVMCLG-ARLI--ADASQCTANGHLNVMVTVYHGLMMFVGVIPALFGGFGNYL 114
 DB 60 EPGIQ--IFHGLQMVYGVGDAALDAGKMFNVFSAHALVMIFWVMPALIGGFANVM 117
 QY 115 MPQIGAPDWAFFRMNLSFWLFIAGTAMGVASLFAFGDGOQSGVGWVLYPPLSTR-E 173
 DB 118 VPLMIGAPDWAFFRMNLSFWLFIAGTAMGVASLFAFGDGOQSGVGWVLYPPLSTR-E 173
 QY 174 AGYSMDLAIFAVHLSGASSIMGAINMTTFLNMRAPGMLHKVLPFSWSIFITAMILLIA 233
 DB 178 PGAVDPALIAHLSGASSILGAINMTTFLNMRAPGMLHKVLPFSWSIFITAMILLIA 237
 QY 234 LPVLGATMLLTDNFGTTFNPPAGGDPILYQHILWFFGHPEVYIIILPGFGIISHVV 293
 DB 238 LPVLGATMLLTDNFGTTFNPPAGGDPILYQHILWFFGHPEVYIIILPGFGIISHVV 297
 QY 294 STSKRPVFCYLPWYAMVAIGVLGVWAHMYTVQMSLTQOSFVMTATNVIAPVTK 353
 DB 298 STSKRPVFCYLPWYAMVAIGVLGVWAHMYTVQMSLTQOSFVMTATNVIAPVTK 357
 QY 354 IFSWIATMGGSVFEXSPMLWAFGMFLFTVGVTGIVLAQGLDRAYHDTYVVAHFY 413
 DB 358 IFSWIATMGGSVFEXSPMLWAFGMFLFTVGVTGIVLAQGLDRAYHDTYVVAHFY 417
 QY 414 VNSLGAIFALFAGIYMPKFSRAPPEWAKLHFMTFFIGANVTFFPOHFLGROGMP 473
 DB 418 VLSGAVFALFAGIYMPKFSRAPPEWAKLHFMTFFIGANVTFFPOHFLGROGMP 477
 QY 474 YIDPAPAFALWNVSSYGAFLAFASLFFIVFVYTL---VAGRRTRPNMGFEADTLE 530
 DB 478 YIDPAPAFALWNVSSYGAFLAFASLFFIVFVYTL---VAGRRTRPNMGFEADTLE 533
 QY 531 WTLPSPPPAHTFTLPK 547
 DB 534 WOLSSPPPPHWEQLPR 550

RESULT 6

S36424
 Cytochrome-c oxidase (EC 1.9.3.1) chain I - Rhizobium leguminosarum
 C;Species: Rhizobium leguminosarum
 C;Date: 09-Dec-1993 #sequence_revision 03-Nov-1995 #text_change 09-Jul-2004
 C;Accession: S36424
 R;Gabel, C.; Bittinger, M.A.; Maier, R.J.
 Submitted to the EMBL Data Library, July 1993
 A;Description: Cytochrome aa3 gene regulation in the Rhizobiaceae: comparison of copper
 A;Reference number: S36424
 A;Accession: S36424
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-538 <GAB>
 A;Cross-references: UNIPROT:Q08855; EMBL:X74341; NID:g396698; PID:g3966
 C;Genetics:
 A;Gene: coxA
 C;Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
 C;Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane-as

rotein

F;31-481/Domain: cytochrome-c oxidase chain I homology <COI>
 F;82-402/Binding site: heme a iron (His) (axial ligands) #status predicted
 F;265-314/Binding site: copper (His) #status predicted
 F;265-269/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
 F;269/Binding site: oxygen (Tyr) #status predicted
 F;392/Binding site: magnesium (His) (shared with chain II) #status predicted
 F;400/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 62.5%; Score 1873.5; DB 2; Length 538;
 Best Local Similarity 64.2%; Pred. No. 2.9e-121;
 Matches 348; Conservative 63; Mismatches 110; Indels 21; Gaps 4;

QY 7 HGHDEHKQGFTHFWSTNHNKIDGLLYLVAAGVGVISLVFTVYMLELMDCGVQMCL 66
 DB 15 HGHDEHAHP---TGWRRSTNHNKIDGLTYLFAITAGTIGGALSIAEAEIYGVYF-- 69
 QY 67 EGARLIADASQCTANGHLNVMVTVYHGLMMFVGVIPALFGGFGNYLMLQIGAPDMAF 126
 DB 70 -----HNTHLNVFVTSHGVIIMFFWMPAMIGGFGNFWFLPMIGAPDMAF 115
 QY 127 PRMNLSFWLFIAGTAMGVASLFAFGDGOQSGVGWVLYPPLSTR-EAGYSMDLAIFAV 185
 DB 116 PRMNLSFWLFIAGTAMGVASLFAFGDGOQSGVGWVLYPPLSTR-EAGYSMDLAIFAV 175
 QY 186 HLCGASSIMGAINMTTFLNMRAPGMLHKVLPFSWSIFITAMILLIAIPVLGATMLL 245
 DB 176 HLCGASSIMGAINMTTFLNMRAPGMLHKVLPFSWSIFITAMILLIAIPVLGATMLL 235
 QY 246 TDRNFGTTFNPPAGGDPILYQHILWFFGHPEVYIIILPGFGIISHVSTFSKRPVGYL 305
 DB 236 TDRNFGTTFNPPAGGDPILYQHILWFFGHPEVYIIILPGFGIISHVSTFSKRPVGYL 295
 QY 306 PMYAMVAIGVLGVWAHMYTVQMSLTQOSFVMTATNVIAPVTKIIFSWIATMGG 365
 DB 296 PMYAMVAIGVLGVWAHMYTVQMSLTQOSFVMTATNVIAPVTKIIFSWIATMGG 355
 QY 366 VEFKSPMLWAFGMFLFTVGVTGIVLAQGLDRAYHDTYVVAHFYVNSLGAIFAIPA 425
 DB 356 VEFKSPMLWAFGMFLFTVGVTGIVLAQGLDRAYHDTYVVAHFYVNSLGAIFAIPA 415
 QY 426 GIYMPKFSRAPPEWAKLHFMTFFIGANVTFFPOHFLGROGMPRIYIDYDEAFALMN 485
 DB 416 GIYMPKFSRAPPEWAKLHFMTFFIGANVTFFPOHFLGROGMPRIYIDYDEAFALMN 475
 QY 486 KVSSYGAFLAFASLFFIVFVYTL---VAGRRTRPNMGFEADTLEWTLPSPPPAHTFT 545
 DB 476 KVSSYGAFLAFASLFFIVFVYTL---VAGRRTRPNMGFEADTLEWTLPSPPPAHTFT 534
 QY 546 PK 547
 DB 535 PR 536

RESULT 7

D87671
 Cytochrome-c oxidase (EC 1.9.3.1), subunit I [imported] - Caulobacter crescentus
 C;Species: Caulobacter crescentus
 C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
 C;Accession: D87671
 R;Niernman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.I.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A;Title: Complete Genome Sequence of Caulobacter crescentus.
 A;Reference number: A87249; MUID:21173698; PMID:11259647
 A;Accession: D87671
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-552 <STO>
 A;Cross-references: UNIPROT:Q9A300; GB:AE005673; NID:g13425118; PIDN:AAK35368.1; GSPDB:G
 C;Genetics:
 A;Gene: CC3406

C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
C:Keywords: copper; electron transfer; heme; iron; magnesium; membrane-associated complex
F:92,411/Binding site: heme a iron (His) (axial ligands) #status predicted
F:274,323,324/Binding site: copper (His) #status predicted
F:274-278/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
F:278/Binding site: oxygen (Tyr) #status predicted
F:401/Binding site: magnesium (His) (shared with chain II) #status predicted
F:409/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 61.6%; Score 1847.5; DB 2; Length 552;
Best Local Similarity 61.6%; Pred. No. 1.8e-119;
Matches 341; Conservative 70; Mismatches 136; Indels 7; Gaps 4;

QY 2 ADAALHGHDKHQGFTRFWSTNTHKDIGLLYLVAAGVGFIISVLFTVYMRLELMDPGV 61
Db 5 ADTHDEAHDHDKHPFFARWFFSTNTHKDIGLLYLVAAGVGFIISVLFTVYMRLELMDPGI 64
QY 62 QYMCLEGARLIADASQCTTANGHLNVMVTVYHGLMFPFVGIPALFGFGNYLPLQIGA 121
Db 65 QVFSDTGML-----AQMGFKGKHGYNVAVTAHALIMIFWVMPAMIGFGFNWFIIMICA 120
QY 122 PDMAFPRMNLISFWLFIAGTANGVASLFAFGDGGQLGSGVGMVLYPPYST-REAGYSMDL 180
Db 121 PDMAFPRMNLISFWLFIAGTANGVASLFAFGDGGQLGSGVGMVLYPPYST-REAGYSMDL 179
QY 181 AIFAVHLSGASSIMGAINMITFLNVRAPGMLTKHVKPLFSWSIFITAWLILLALPVLGA 240
Db 180 ALLSLHLGASSILGAINFITTFNRRAPGMLTKHVKPLFSWSIFITAWLILLALPVLGA 239
QY 241 ITMLLTDNRFGTFFNPPAGGDPILYQHILMFPFGHPEVYIILPGFIIISHVSTFSKKP 300
Db 240 ITMLLTDNRFGTFFNPPAGGDPILYQHILMFPFGHPEVYIILPGFIIISHVSTFSKKP 299
QY 301 VEGYLPMYAMVAIGVLGFWVAHMYTVGMSLTQGSFPMATWVIAVPTGKIFSWIAT 360
Db 300 VEGYLAMAYAMVAIGVGFVVAHMYTVGMSLTQGSFPMATWVIAVPTGKIFSWIAT 359
QY 361 MWGGSVEFKPSMLWAFGMFLFTVGGVTGIVLAQAGLDRVHDTYVVAHFHYVMSLCAI 420
Db 360 MWGGSLSFKTMLWAFGFIIFLTVGGVTGIVLSNAGIDYTLHDTYVVAHFHYVLSLCAV 419
QY 421 FAIFAGIYFPMKFGSRAPPEWAALKHFWTPIGANTFTFPOHFLGRCMPRRYIDYDEA 480
Db 420 FAIFAGIYFPMKFGSRAPPEWAALKHFWTPIGANTFTFPOHFLGRCMPRRYIDYDEA 479
QY 481 FALWKNVSSYGAFLAFASLFLFIVFVYTLVAGRETRPNMGPEADTLEWTLSPPPAH 540
Db 480 FLMWYVSTVGWMTIVVGCVFNLVLEIAAIR-RRAEANPWEGATILEWTLSPPPFFH 538
QY 541 TFEETLPRKSDWDKH 554
Db 539 QSEPPVKGDDHH 552

RESULT 8
00ZJ1
C:Species: Bradyrhizobium japonicum
C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 09-Jul-2004
C:Accession: S13076; S12101
R:Botz, M.; Bolliger, M.; Hennecke, H.
Mol. Microbiol. 4, 2147-2157, 1990
A:Title: Genetic analysis of the cytochrome c-aa(3) branch of the Bradyrhizobium japonicum
A:Reference number: S13076; MUID:91211625; PMID:1965217
A:Accession: S13076
A:Molecule type: DNA
A:Residues: 1-541 <BOT>
A:Cross-references: UNIPROT:P31933; EMBL:X54800; NID:G39502; PIDN:CAA38570.1; PID:G39503
R:Gabel, C.; Maier, R.U.
Nucleic Acids Res. 18, 6143, 1990
A:Title: Nucleotide sequence of the cox a gene encoding subunit I of cytochrome aa(3) of
A:Reference number: S12101; MUID:91045095; PMID:2172930
A:Accession: S12101

A:Molecule type: DNA
A:Residues: 1-541 <GAB>
A:Cross-references: EMBL:X54318; NID:G39505; PIDN:CAA38216.1; PID:G39506
C:Genetics:
A:Gene: coxa
C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
C:Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane-associated protein
F:34-484/Domain: cytochrome-c oxidase chain I homology <CO1>
F:41-60/Domain: transmembrane #status predicted <TM01>
F:87-107/Domain: transmembrane #status predicted <TM02>
F:124-142/Domain: transmembrane #status predicted <TM03>
F:173-193/Domain: transmembrane #status predicted <TM04>
F:211-231/Domain: transmembrane #status predicted <TM05>
F:262-282/Domain: transmembrane #status predicted <TM06>
F:294-315/Domain: transmembrane #status predicted <TM07>
F:331-352/Domain: transmembrane #status predicted <TM08>
F:365-385/Domain: transmembrane #status predicted <TM09>
F:404-424/Domain: transmembrane #status predicted <TM10>
F:441-460/Domain: transmembrane #status predicted <TM11>
F:483-503/Domain: transmembrane #status predicted <TM12>
F:85,405/Binding site: heme a iron (His) (axial ligands) #status predicted
F:268,317,318/Binding site: copper (His) #status predicted
F:268-272/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
F:272/Binding site: oxygen (Tyr) #status predicted
F:403/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 61.0%; Score 1829; DB 1; Length 541;
Best Local Similarity 61.8%; Pred. No. 3.3e-118;
Matches 337; Conservative 72; Mismatches 112; Indels 24; Gaps 4;

QY 7 HGHDKHQGFTRFWSTNTHKDIGLLYLVAAGVGFIISVLFTVYMRLELMDPGVQVACL 66
Db 15 HGHDEHAHPTGWRYYVSTNTHKDIGTMYLIFAVIAGVIGAAVSIAIAELMYQVQIF-- 72
QY 67 EGARLIADASQCTTANGHLNVMVTVYHGLMFPFVGIPALFGFGNYLPLQAGPDMAF 126
Db 73 -----HETHYTVFVTSGLIMIFWVMPAMIGFGFNWFIIMIGAPDMAF 118
QY 127 PRMNLISFWLFIAGTANGVASLFAFGDGGQLGSGVGMVLYPPYST-REAGYSMDLAFV 185
Db 119 PRMNLISFWLFIAGTANGVASLFAFGDGGQLGSGVGMVLYPPYST-REAGYSMDLAFV 178
QY 186 HLGASSIMGAINMITFLNVRAPGMLTKHVKPLFSWSIFITAWLILLALPVLGAITMLL 245
Db 179 HLGASSILGAINFITTFNRRAPGMLTKHVKPLFSWSILVTVFVLSLFLPVLGAITMLL 238
QY 246 TDRNFGTFFNPPAGGDPILYQHILMFPFGHPEVYIILPGFIIISHVSTFSKKPVGYL 305
Db 239 TDRNFGTFFNPPAGGDPILYQHILMFPFGHPEVYIILPGFIIISHVSTFSKKPVGYL 298
QY 306 PMYAMVAIGVLGFWVAHMYTVGMSLTQGSFPMATWVIAVPTGKIFSWIATMWGGS 365
Db 299 GMAYAMVAIGGIGFVVAHMYTVGMSLTQGSFPMATWVIAVPTGKIFSWIATMWGGS 358
QY 366 VEFKSPMLWAFGMFLFTVGGVTGIVLAQAGLDRVHDTYVVAHFHYVMSLCAIFAIPA 425
Db 359 IEFAPRMIWAGFIIFLTVGGVTGIVLANAGVDRVLEQETYYVVAHFHYVLSLCAIFAIPA 418
QY 426 GIYFPMKFGSRAPPEWAALKHFWTPIGANTFTFPOHFLGRCMPRRYIDYDEAPALWN 485
Db 419 GWYTFPMTGYMYNETLAKAHFWTPIGANTFTFPOHFLGRCMPRRYIDYDEAPAGWN 478
QY 486 KVSSYGAFLAFASLFLFIVFVYTLVAGRETRPNMGPEADTLEWTLSPPPAHPT 542
Db 479 LVSSVGSYISG-----FGVLIFLYCVIDAFKVKVAGDNPWAGATILEWTLSPPPPHQF 534
QY 543 ETLPK 547
Db 535 EVLPR 539

RESULT 9

S62706
 Cytochrome-c oxidase (EC 1.9.3.1) chain I - green alga (Platymonas subcordiformis) mitochondrion
 C:Species: mitochondrion Platymonas subcordiformis
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
 C:Accession: S62706
 R:Kessler, U.; Zetsche, K.
 Plant Mol. Biol. 29, 1081-1086, 1995
 A:Title: Physical map and gene organization of the mitochondrial genome from the unicell
 A:Reference number: S62702; MUID:96145517; PMID:8555450
 A:Accession: S62706
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-523 <RES>
 A:Cross-references: UNIPROT:Q36520; EMBL:Z47795; NID:9633584; PIDN:CRA87753.1; PID:96335
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1995
 C:Genetics:
 A:Gene: coi
 A:Genome: mitochondrion
 C:Function:
 A:Description: the cytochrome-c oxidase complex catalyzes the oxidation of four molecules
 ns from the mitochondrial matrix producing two molecules of water and lowering the conce
 A:Pathway: oxidative phosphorylation; respiratory chain
 A:Note: chain I directly reduces oxygen on the mitochondrial matrix side of the inner-me
 C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
 C:Keywords: cytochrome; copper; electron transfer; heme; iron; lipoprotein; magnesium
 A:active phosphorylation; oxidoreductase; respiratory chain; transmembrane protein
 F:1-13/Domain: mitochondrial matrix #status predicted <MM1>
 F:13-460/Domain: cytochrome-c oxidase chain I homology <CO1>
 F:13-460/Domain: transmembrane #status predicted <TM01>
 F:14-42/Domain: intracristal #status predicted <ITC1>
 F:43-54/Domain: transmembrane #status predicted <TM02>
 F:55-90/Domain: transmembrane #status predicted <TM02>
 F:91-98/Domain: mitochondrial matrix #status predicted <MM2>
 F:99-121/Domain: transmembrane #status predicted <TM03>
 F:122-144/Domain: intracristal #status predicted <ITC2>
 F:145-174/Domain: transmembrane #status predicted <TM04>
 F:175-186/Domain: mitochondrial matrix #status predicted <MM3>
 F:187-216/Domain: transmembrane #status predicted <TM05>
 F:217-231/Domain: intracristal #status predicted <ITC3>
 F:232-265/Domain: transmembrane #status predicted <TM06>
 F:266-272/Domain: mitochondrial matrix #status predicted <MM4>
 F:273-289/Domain: transmembrane #status predicted <TM07>
 F:290-301/Domain: intracristal #status predicted <ITC4>
 F:302-330/Domain: transmembrane #status predicted <TM08>
 F:331-338/Domain: mitochondrial matrix #status predicted <MM5>
 F:339-359/Domain: transmembrane #status predicted <TM09>
 F:360-373/Domain: intracristal #status predicted <ITC5>
 F:374-403/Domain: transmembrane #status predicted <TM10>
 F:404-409/Domain: mitochondrial matrix #status predicted <MM6>
 F:410-436/Domain: transmembrane #status predicted <TM11>
 F:437-449/Domain: intracristal #status predicted <ITC6>
 F:450-481/Domain: transmembrane #status predicted <TM12>
 F:482-522/Domain: mitochondrial matrix #status predicted <MM7>
 F:65,381/Binding site: heme a iron (His) (axial ligands) #status predicted
 F:244,293,294/Binding site: copper (His) #status predicted
 F:244,293,294/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
 F:248/Binding site: oxygen (Tyr) #status predicted
 F:322/Binding site: myristate (Lys) (covalent) #status predicted
 F:371/Binding site: magnesium (His) (shared with chain II) #status predicted
 F:379/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 59.1%; Score 1773; DB 2; Length 523;
 Best Local Similarity 60.1%; Pred. No. 2,2e-114;
 Matches 322; Conservative 82; Mismatches 104; Indels 28; Gaps 6;

QY 17 FFFRWFSTNHHKDIGLLYLVAAGVGFISVLFTVYWRLELMDPGVQVYMCLEGARLIADAS 76
 DB 4 FAQRLWLFSTNHHKDIGLLYLVAAGVGFISVLFTVYWRLELMDPGVQVYMCLEGARLIADAS 48
 QY 77 QTCTANGHLNWNVYTHGILMMFFVGIPLFGFGNGYLMPLQIGAPDMAPPRNNLSFWL 136
 DB 49 QVLGNHQLNVLITAHAFIMFPMVWPAJIGFGNGVFPIMIGAPDMAPPRNNLSFWL 108
 QY 137 FIAGTANGVASLFPAGGGDGLGSGVGVLYPPPLST--REAGYSMDLAIFAVHLSGASSIM 194

Db 109 LPPSLLLSALV-----FAGAGTGVTVYPPSIASHSGASVDLAFSLHSGASSIL 163
 QY 195 GAINMTTFLNMRAPGNTLHKVPLFSWISFITAWLILLALPVLAGAITMLTDRNFQTTTF 254
 Db 164 CALNFITITLNMFGMTHRLPLFWMSVLITAPLALLSLPLVLAGAITMLTDRNFQTTTF 223
 QY 255 FNPAGGGDPLLYOHLNFFGHEPEVYIILFGFGLIHSVSTFESKPKVGVGLPVMYAVAI 314
 Db 224 FDPAGGGDPLILFQHLFFGHEPEVYIILFAGFGLIHSVSTFESKPKVGVGLVMYAMLSI 283
 QY 315 GVLGVVWAHMYTVGMSLTQQSYFMILATVIAVPTGKIFSWIATMWGGSVFESKPSMLW 374
 Db 284 GILGFIVWAHMYVVGDLIDITRAYFTAAITLIIAVPTGKIFSWIATMWGGSIEFRTPMLF 343
 QY 375 AFGWMLFTVGGVTVGIVLAQGLDRAYHDTYVVAHMYHYSIGALPAITAGIIFYMPKF 434
 Db 344 AIGFLFTVGGVTVGVVLANSGDLIALHDTYVVAHMYHYSIGALPAITAGIIFYMGIKI 403
 QY 435 SGRAPPEWAALHETWTFIIGANVTFFQHLGRCMPRRVIDYPEAFALMNKYSYSGAFL 494
 Db 404 CGLQYSEFELGQIHWLFFVGVNITFFPMHFLGSGMPRRIPDYPDAFGWNAVCSYGL 463
 QY 495 AFASLFFIVFVYTLVAGRRTRPNW---GEFADLTLEWTLPSPPPAHTFTFLP 546
 Db 464 SVLGALFFVYV-VYDTLSGBERCPKNWETVPQTSA-TLEWLTSPSPPAHTFTFEVP 517
 RESULT 10
 S25956
 Cytochrome-c oxidase (EC 1.9.3.1) chain I - liverwort (Marchantia polymorpha) mitochondrion
 C:Species: mitochondrion Marchantia polymorpha
 C:Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 09-Jul-2004
 C:Accession: S25956
 R:Oda, K.; Yamato, K.; Ohta, E.; Nakamura, Y.; Takemura, M.; Nozato, N.; Akashi, K.; Kane
 J. Mol. Biol. 223, 1-7, 1992
 A:Title: Gene organization deduced from the complete sequence of liverwort Marchantia pol
 A:Reference number: S25941; MUID:92114051; PMID:1731062
 A:Accession: S25956
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-522 <ODA>
 A:Cross-references: UNIPROT:P26856; EMBL:M68929; NID:9786182; PIDN:AAC09451.1; PID:97862
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992
 C:Genetics:
 A:Gene: coi
 A:Genome: mitochondrion
 A:Introns: 15/2; 60/1; 125/3; 132/2; 171/1; 208/3; 243/3; 372/3; 435/3
 C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
 C:Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane-as
 transmembrane protein
 F:13-460/Domain: cytochrome-c oxidase chain I homology <CO1>
 F:65,381/Binding site: heme a iron (His) (axial ligands) #status predicted
 F:244,293,294/Binding site: copper (His) #status predicted
 F:244,293,294/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
 F:248/Binding site: oxygen (Tyr) #status predicted
 F:371/Binding site: magnesium (His) (shared with chain II) #status predicted
 F:379/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 58.6%; Score 1756.5; DB 2; Length 522;
 Best Local Similarity 58.9%; Pred. No. 3e-113;
 Matches 315; Conservative 92; Mismatches 101; Indels 27; Gaps 6;

QY 17 FFFRWFSTNHHKDIGLLYLVAAGVGFISVLFTVYWRLELMDPGVQVYMCLEGARLIADAS 76
 Db 4 FAQRLWLFSTNHHKDIGLLYLVAAGVGFISVLFTVYWRLELMDPGVQVYMCLEGARLIADAS 48
 QY 77 QTCTANGHLNWNVYTHGILMMFFVGIPLFGFGNGYLMPLQIGAPDMAPPRNNLSFWL 136
 Db 49 QVLGNHQLNVLITAHAFIMFPMVWPAJIGFGNGVFPIMIGAPDMAPPRNNLSFWL 108
 QY 137 FIAGTANGVASLFPAGGGDGLGSGVGVLYPPPLS--TREAGYSMDLAIFAVHLSGASSIM 194


```
Db 109 LPPSLLLSSALV-----EVGCGSGWTVPPLSGITSHSGGSDVLAIFSLHLSGVSSIL 163
Qy 195 GAINMITFLNMRAGMTLHKVPLFSWGFITAMILLALLPVLAGAITMLLTDNFGTTF 254
Db 164 GSNITITFENRPAFLMHRLEPLFVNSVLVTAFLLLSLPVLAGAITMLLTDNRFNFTF 223
Qy 255 FNPAGGDPILYQHILWFFGHPEVYIILPQFGIISHVVSTFSKPKVFGYLPYAMVAI 314
Db 224 FDPAGGDPILYQHILWFFGHPEVYIILPQFGIISHIVSTFSRKPVPFGYLGWYAMISI 283
Qy 315 GVLGFVWAHMYTVMGSLTQOSYFMLATMVIATVGTGKISWTATMGGSVKPSMLW 374
Db 284 GVLGFVWAHMYTVMGSLTQOSYFMLATMVIATVGTGKISWTATMGGSIQYKTMPLF 343
Qy 375 AFGFMFLFTVGGVGTGIVLAQAGLDRAVDHTYVVAHFHYVMSLGAIFAIFAIFAGIYFMPKF 434
Db 344 AVGFIFLFTVGGVGTGIVLAQAGLDRAVDHTYVVAHFHYVMSLGAIFAIFAIFAGIYFMPKF 403
Qy 435 SGRAPEWAUKHFWTFFIGANVTFFPOHFLGRQGNPRYIDYPAFALMKNVSSYGAFL 494
Db 404 TGLQYPETLQGHFWITFFGNLTFFPMHFLGLAGMPRRIPDYDAYAGWNAFSSFGSYV 463
Qy 495 AFAS-FLRFIVFVTLVAGRETRPNWG--EFADTLEWTLPSPPPAHTPETLP 546
Db 464 SVVGICFVTVVL--TLTSENKCAPSPWAVEQNSTTLEWVPSPPAHTPETLP 516
RESULT 11
cytochrome-c oxidase (EC 1.9.3.1) - Rickettsia conorii (strain Malieh 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C:Accession: A97769
R:Ogata, H.; Audic, S.; Reneato-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Rickettsia conorii
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: A97769
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-532 <KUR>
A:Cross-references: UNIPROT:Q92167; GB:AE006914; PIDN:AAL03091.1; PID:G15619633; GSPDB:G15619633
A:Gene: coxa
C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
C:Keywords: copper; electron transfer; heme; iron; magnesium; membrane-associated complex
F:79/395/Binding site: heme a iron (His) (axial ligands) #status predicted
F:258/307/Binding site: copper (His) #status predicted
F:258-262/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
F:262/Binding site: oxygen (Tyr) #status predicted
F:385/Binding site: magnesium (His) (shared with chain II) #status predicted
F:393/Binding site: heme a3 iron (His) (axial ligand) #status predicted
Query Match 58.6%; Score 1756.5; DB 2; Length 532;
Best Local Similarity 59.3%; Pred. No. 3.1e-113;
Matches 320; Conservative 78; Mismatches 119; Indels 23; Gaps 4;
Qy 9 HDHHEKQGFTEWSTNKHDKGLAYLVAAGVVGFISVLFTVYMRLEMDPQVQVWCLG 68
Db 10 HDGHTPHGWRRLSTNKHDKGLAYLVAAGVVGFISVLFTVYMRLEMDPQVQVWCLG 65
Qy 69 ARLIADASOTCANGHLNMYVTHGILMMFFVGIPALFGGFGNLYLQIGAPDMAFPR 128
Db 66 -----NHDFOLYNLVITAHAVIMVFMWMPALFGGFGNLYLQIGAPDMAFPR 114
Qy 129 MNLSFWLFIAGTAMGVASLFPAGDGGQLGSGVGVWLYPPLS--TREAGYSMDLAIFAVH 186
Db 115 LNNISFWLVPAPFLMGSAFVDG-----GPGTGTWLYPPLSLNLSGHCFAAVDMAIFSLH 169
Qy 187 LSGASSIMGAINMITFLNMRAGMTLHKVPLFSWGFITAMILLALLPVLAGAITMLLT 246
Db 170 L'GLSILGSLINLIVITFNMRAPGMLPKMPLFVNSILVTAFLIILAMPVIGGATMLLT 229
```

```
Qy 247 DRNFGTTFNPAGGDPILYQHILWFFGHPEVYIILPQFGIISHVVSTFSKPKVFGYLP 306
Db 230 DRNFGTTFNPAGGDPILYQHILWFFGHPEVYIILPQFGIISHVVSTFSKPKVFGYLP 289
Qy 307 MYAMVAIGVLGFVWAHMYTVMGSLTQOSYFMLATMVIATVGTGKISWTATMGGSV 366
Db 290 MIGAMVIIGVGFVWAHMYTVMGSLTQOSYFMLATMVIATVGTGKISWTATMGGSI 349
Qy 367 BFKSPMLWAFGPMFLFTVGGVGTGIVLAQAGLDRAVDHTYVVAHFHYVMSLGAIFAIFAG 426
Db 350 TFFTPMLFSGFILLFTIGGVTGIIISNSALDVLHDYVVAHFHYTWSLGLATFAPAG 409
Qy 427 IYFMPKESGRAPPEWAUKHFWTFFIGANVTFFPOHFLGRQGNPRYIDYPAFALMKN 486
Db 410 FYWFGKISQKQPDILGKTHFWITFGVNLTFPQHFHFLGLAGMPRRIPDYPAFAGWNN 469
Qy 487 VSSVGAFLAFLFATFIFVITLVAGRETRPNWGFEADTLEWTLPSPPPAHTPETLP 546
Db 470 VSSIGAGISWFAALYFVIVFYTLKYG-KDCFPNPMGEGADTLEWTLSPPPHTPETTP 528
RESULT 12
S14138
cytochrome-c oxidase (EC 1.9.3.1) chain I - sugar beet mitochondrion
C:Species: mitochondrion Beta vulgaris var. altissima (sugar beet)
C:Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 09-Jul-2004
C:Accession: S14138
R:Senda, M.; Harada, T.; Mikami, T.; Sugitara, M.; Kinoshita, T.
Curr. Genet. 19, 175-181, 1991
A:Title: Genomic organization and sequence analysis of the cytochrome oxidase subunit II
A:Reference number: S14138; MUID:91330331; PMID:1651175
A:Accession: S14138
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-524 <SEN>
A:Cross-references: UNIPROT:P24794; GB:X57693; GB:S47702; NID:g11258; PIDN:CAA40874.1; PJ
C:Genetics:
A:Gene: coxi
A:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
C:Keywords: chromoprotein; copper; electron transfer; heme; iron; magnesium; membrane-associated protein
F:12-459/Domain: cytochrome-c oxidase chain I homology <CO1>
F:64/380/Binding site: heme a iron (His) (axial ligands) #status predicted
F:243/292/Binding site: copper (His) #status predicted
F:243-247/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
F:247/Binding site: oxygen (Tyr) #status predicted
F:370/Binding site: magnesium (His) (shared with chain II) #status predicted
F:378/Binding site: heme a3 iron (His) (axial ligand) #status predicted
Query Match 58.3%; Score 1748.5; DB 2; Length 524;
Best Local Similarity 59.2%; Pred. No. 1.1e-112;
Matches 315; Conservative 86; Mismatches 106; Indels 25; Gaps 5;
Qy 20 RWFMTNKHDKGLAYLVAAGVVGFISVLFTVYMRLEMDPQVQVWCLGADIASQTC 79
Db 6 RULPSTNKHDKGLAYLVAAGVVGFISVLFTVYMRLEMDPQVQVWCLGADIASQTC 50
Qy 80 TANGHLNMYVTHGILMMFFVGIPALFGGFGNLYLQIGAPDMAFPRMNLSEWLFIA 139
Db 51 GGNHQLYNLVITAHAVIMVFMWMPALFGGFGNLYLQIGAPDMAFPRMNLSEWLFIA 110
Qy 140 GTAMGVASLFPAGDGGQLGSGVGVWLYPPLS--TREAGYSMDLAIFAVHLSGASSIMGAI 197
Db 111 SLULLSSALV-----EVGSGTGTWLYPPLSGITSHSGGAVDLAIFSLHLSGVSSILGI 165
Qy 198 NMITFLNMRAGMTLHKVPLFSWGFITAMILLALLPVLAGAITMLLTDNFGTTFNPN 257
Db 166 NFITTFENMRAGMTLHKVPLFSWGFITAMILLALLPVLAGAITMLLTDNFGTTFNPN 225
Qy 258 AGGSDPILYQHILWFFGHPEVYIILPQFGIISHVVSTFSKPKVFGYLPYAMVAIGVL 317
Db 226 AGGSDPILYQHILWFFGHPEVYIILPQFGIISHVVSTFSKPKVFGYLPYAMVAIGVL 285
```

```

QY 318 GFVVAAHMYTVGMSLTQOOSYFVLATWIAVPTGKIFSWIATMMGSGVEKSPMLWAPG 377
Db 286 GFLVAAHMYTVGMSLTQOOSYFVLATWIAVPTGKIFSWIATMMGSGIQYTPMLFVAG 345
QY 378 FMFLFTVGGVTGVLAAQAGLDRAHYDHTYVVAHFHYVMSLGAIFAIFAGIYFMPKFSGR 437
Db 346 FIEFLTGGTGLTVLANSGLDIALHDTYVVAHFHYVMSLGAIFAIFAGIYFMPKFSGR 405
QY 438 AFPEWAAKLFHFWFFIGANVTFFPQHLGRQGMPPRYIDYPEAFALWNKVSYGAFAPLA 497
Db 406 TYPETLQIHFWLTFPGVNTFFPQHLGRQGMPPRYIDYPEAFALWNKVSYGAFAPLA 465
QY 498 SFLEFFIVFYVTLVAGR-RETENPAG--EFADTLEWTLPSPPPAHTFTLP 546
Db 466 GICCFVFWVITLSSGKNKRCAPSVAWEENSTLEWMLPSPAPHTFGELP 517

RESULT 13
T11913
cytochrome-c oxidase (EC 1.9.3.1) chain I - Prototheca wickerhamii mitochondrion
C:Species: mitochondrion Prototheca wickerhamii
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C:Accession: T11913
R:Wolff, G.; Plante, I.; Lang, B.F.; Kueck, U.; Burger, G.
J. Mol. Biol. 237, 75-86, 1994
A:Title: Complete sequence of the mitochondrial DNA of the chlorophyte alga Prototheca
A:Reference number: Z17373; MUID:94180393; PMID:8133522
A:Accession: T11913
A:Status: preliminary; translated from GB/EMBL/DDBB
A:Molecule type: DNA
A:Residues: 1-515 <MOL>
A:Cross-references: UNIPROT:Q37616; EMBL:U02970; NID:g467843; PID:g467847; PIDN:AAD12634
A:Experimental source: strain 263-11
C:Genetics:
A:Genome: mitochondrion
A:Introns: 129/2; 237/1; 240/3
A:Note: cox1
C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
C:Keywords: copper; electron transfer; heme; iron; magnesium; membrane-associated complex
F:10-457/Domain: cytochrome-c oxidase chain I homology <COI>
F:62,378/Binding site: heme a iron (His) (axial ligands) #status predicted
F:241,290,291/Binding site: copper (His) #status predicted
F:241-245/Cross-link: 1'-histidyl-3'-tyrosine (His-Tyr) #status predicted
F:245/Binding site: oxygen (Tyr) #status predicted
F:368/Binding site: magnesium (His) (shared with chain II) #status predicted
F:376/Binding site: heme a3 iron (His) (axial ligand) #status predicted

Query Match 58.2%; Score 1745; DB 2; Length 515;
Best Local Similarity 58.2%; Pred. No. 1.8e-112;
Matches 310; Conservative 92; Mismatches 103; Indels 28; Gaps 6;

QY 19 TRFMSNTHKDIGLLYLVAAGVGVTSVLTFTVMRLMLMDPQVQVYVCLSGARLIADASQT 78
Db 3 TRWLYSTNTHKDIGTMYLIFGAFSGVLTGTFSLIRMLAQPG-----N01 47
QY 79 CTANGHLNMYVTHYHGLMMFFGIPALFGGNGYLMPLQIGAPDMAPFRMNNLSFWLPI 138
Db 48 LNHQHLNLYVITAHAFPLMFFMLMPALMGFGNFWLPIILIGAPDMAPFRMNNLSFWLPI 107
QY 139 AGTAMGVASLPAFGDQGLGSGVWVLYPLST--REAGYSMDLAFPAVHLGASISGA 196
Db 108 PSLLLLVSSALV-----EVGAGTGWTPYPLASIASHSGSGVDLAFSLHLAGVSSILGA 162
QY 197 INMITFLNMRAPGWLTHKVPFLPSWSIFITAWLILLALPVLAGAIFMLLTDNRNFGTTFN 256
Db 163 INICTVFNMRAPGMSHRLPLFVNAVITAWLILLALCLPLVLAGITMLLTDNRNFTSFD 222
QY 257 PAGGDPILYQHILWFFGHPHVVIIILPGGIIISHVVSFTSKFPVGYLPMYAMVAIGV 316
Db 223 PAGGDPILYQHILWFFGHPHVVIIILPGGIIISHVVSFTSKFPVGYLPMYAMCSIGI 282
QY 317 LGFVAAHMYTVGMSLTQOOSYFVLATWIAVPTGKIFSWIATMMGSGVEKSPMLWAP 376

```

```

Db 283 LGFVAAHMYTVGMSLTQOOSYFVLATWIAVPTGKIFSWIATMMGSGTELETPMLFAV 342
QY 377 GFMEFLTGGVTGVLAAQAGLDRAHYDHTYVVAHFHYVMSLGAIFAIFAGIYFMPKFSG 436
Db 343 GFLFLTGGTGLTVLANSGLDVAHDTYVVAHFHYVMSLGAIFAIFAGIYFMPKFSG 402
QY 437 RAFFEWAAKLFHFWFFIGANVTFFPQHLGRQGMPPRYIDYPEAFALWNKVSYGAFAPLA 496
Db 403 LQYETLQIHFWLTFPGVNTFFPQHLGRQGMPPRYIDYPEAFALWNKVSYGAFAPLA 462
QY 497 ASFLFFIVFYVTLVAGR-RETENPAG--EFADTLEWTLPSPPPAHTFTLP 545
Db 463 TAVLFFVYVYKTLTS--NEVCPNPNWETTPGVSPTLEWMLPSPAPHTFEEI 513

RESULT 14
S65346
cytochrome-c oxidase (EC 1.9.3.1) chain I - tomato mitochondrion
C:Species: mitochondrion Lycopersicon esculentum (tomato)
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C:Accession: S65346; S65080
R:Kadowaki, K.; Ozawa, K.; Kazama, S.; Kubo, N.; Akihama, T.
Curr. Genet. 28, 415-422, 1995
A:Title: Creation of an initiation codon by RNA editing in the coxI transcript from tomato
A:Reference number: S65080; MUID:96155620; PMID:8575013
A:Accession: S65346
A:Molecule type: mRNA
A:Residues: 1-527 <KAD>
A:Cross-references: UNIPROT:Q34882; EMBL:X54738
A:Experimental source: strain Ponderosa; leaf
A:Note: 1-Wet, 4-Leu, 85-Phe, 151-Phe, 184-Leu, 197-Leu, 223-Phe, 249-Leu, 254-Phe, 395-7
A:Accession: S65080
A:Molecule type: DNA
A:Residues: 'T', 2-3, 'P', 5-84, 'S', 152-183, 'S', 185-196, 'P', 198-222, 'S', 224-248, '
A:Cross-references: EMBL:X54738; NID:g1165117; PIDN:CAA38535.1; PID:g1165118
A:Experimental source: strain Ponderosa; leaf
C:Genetics:
A:Gene: coxI
A:Genome: mitochondrion
C:Function:
A:Description: the cytochrome-c oxidase complex catalyzes the oxidation of four molecules
ns from the mitochondrial matrix producing two molecules of water and lowering the concen
A:Pathway: oxidative phosphorylation; respiratory chain
A:Note: chain I directly reduces oxygen on the mitochondrial matrix side of the inner-mem
C:Superfamily: cytochrome-c oxidase chain I; cytochrome-c oxidase chain I homology
C:Keywords: chromoprotein; copper; electron transfer; heme; iron; lipoprotein; magnesium;
active phosphorylation; oxidoreductase; respiratory chain; RNA editing; transmembrane prot
F:1-12/Domain: mitochondrion matrix #status predicted <MM1>
F:12-459/Domain: cytochrome-c oxidase chain I homology <COI>
F:13-41/Domain: transmembrane #status predicted <TM01>
F:42-53/Domain: intracristal #status predicted <ITC1>
F:54-89/Domain: transmembrane #status predicted <TM02>
F:90-97/Domain: mitochondrion matrix #status predicted <MM2>
F:98-120/Domain: transmembrane #status predicted <TM03>
F:121-143/Domain: intracristal #status predicted <ITC2>
F:144-173/Domain: transmembrane #status predicted <TM04>
F:174-185/Domain: mitochondrion matrix #status predicted <MM3>
F:186-215/Domain: transmembrane #status predicted <TM05>
F:216-230/Domain: intracristal #status predicted <ITC3>
F:231-264/Domain: transmembrane #status predicted <TM06>
F:265-271/Domain: mitochondrion matrix #status predicted <MM4>
F:272-288/Domain: transmembrane #status predicted <TM07>
F:289-300/Domain: intracristal #status predicted <ITC4>
F:301-329/Domain: transmembrane #status predicted <TM08>
F:330-337/Domain: mitochondrion matrix #status predicted <MM5>
F:338-358/Domain: transmembrane #status predicted <TM09>
F:359-372/Domain: intracristal #status predicted <ITC5>
F:373-402/Domain: transmembrane #status predicted <TM10>
F:403-408/Domain: mitochondrion matrix #status predicted <MM6>
F:409-435/Domain: transmembrane #status predicted <TM11>
F:436-448/Domain: intracristal #status predicted <ITC6>
F:449-480/Domain: transmembrane #status predicted <TM12>

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2004, 23:28:06 ; Search time 197 Seconds
(without alignments)

1626.821 Million cell updates/sec

Title: US-09-712-768A-2

Perfect score: 3000

Sequence: 1 MADAALHGHDHHEKQGFTR.....PAHIFETLPKRSDDWKHPSH 557

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_02:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
1	2538.5	84.6	558	1	CX1B_PARDE	P98002 paracoccus
2	2493.5	83.1	566	1	CX1B_RHOSH	P33517 rhodobacter
3	2347	78.2	554	1	CX1A_PARDE	P08305 paracoccus
4	1898.5	63.3	562	2	Q92RG9	Q92RG9 rhizobium m
5	1897.5	63.2	572	2	Q8UHB4	Q8UHB4 agrobacteri
6	1891.5	63.0	552	2	Q8Q263	Q8Q263 brucella su
7	1890.5	63.0	552	2	Q8TFQ5	Q8TFQ5 brucella me
8	1873.5	62.5	538	1	CX1I_RHILE	Q08855 rhizobium l
9	1847.5	61.6	552	2	Q9A300	Q9A300 caulobacter
10	1846.5	61.6	550	2	Q985X1	Q985X1 rhizobium l
11	1842	61.4	541	2	Q6NBJ6	Q6NBJ6 rhodospseudo
12	1842	61.4	541	2	CAE26276	CAE26276 rhodopseu
13	1838	61.3	539	2	Q39631	Q39631 nitrobacter
14	1829	61.0	541	1	CX1I_BRAJA	P31833 bradyrhizob
15	1773	59.1	523	2	Q36520	Q36520 platymonas
16	1762.5	58.6	538	2	Q8W9R3	Q8W9R3 mesostigma
17	1756.5	58.6	522	1	CX1I_MARPO	P26856 marchantia
18	1756.5	58.6	525	2	Q9TCA6	Q9TCA6 nectroselimi
19	1756.5	58.6	532	2	Q92167	Q92167 rickettsia
20	1751.5	58.4	525	2	Q63228	Q63228 populus tre
21	1748.5	58.3	524	1	CX1I_BETVU	P24794 beta vulgar
22	1748.5	58.3	524	2	Q7HJM3	Q7HJM3 beta vulgar
23	1748.5	58.3	524	2	BAA86630	BAA86630 beta vulg
24	1748.5	58.3	524	2	BAA99455	BAA99455 beta vulg
25	1748.5	58.3	532	2	Q7PBR5	Q7PBR5 rickettsia
26	1746.5	58.2	526	2	Q8M1D5	Q8M1D5 chaetosphae
27	1745	58.2	515	2	Q37616	Q37616 prototheca
28	1744	58.1	534	1	CX1I_RICPR	Q54069 rickettsia
29	1739.5	58.0	524	2	Q7YAN3	Q7YAN3 chara vulga
30	1738.5	58.0	527	1	CX1I_ARATH	P60620 arabidopsis
31	1738.5	58.0	527	1	CX1I_RAPSA	P60621 rapanus sa

RESULT 1

ID CX1B_PARDE STANDARD; PRT; 558 AA.

AC P98002;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-OCT-2004 (Rel. 45, Last annotation update)

DE Cytochrome c oxidase polypeptide I-beta (EC 1.9.3.1) (Cytochrome aa3

subunit I-beta).

GN Name=ctadII;

OS Paracoccus denitrificans.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacteriales;

OC Rhodobacteraceae; Paracoccus.

OX NCBI_TaxID=266;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PD 1222;

RX MEDLINE=90184495; PubMed=2155830;

RA Raitio M., Pispas J.M., Metso T., Saraste M.;

RT "Are there isoenzymes of cytochrome c oxidase in Paracoccus

denitrificans?";

RL FEBS Lett. 261:431-435(1990).

RN [2]

RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).

RX MEDLINE=95379947; PubMed=7651515;

RA Iwata S., Ostermeier C., Ludwig B., Michel H.;

RT "Structure at 2.8 A resolution of cytochrome c oxidase from Paracoccus

denitrificans.";

RL Nature 376:660-669(1995).

RN [3]

RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).

RX MEDLINE=98021406; PubMed=9380672;

RA Ostermeier C., Harrenga A., Ermler U., Michel H.;

RT "Structure at 2.7-A resolution of the Paracoccus denitrificans two-

subunit cytochrome c oxidase complexed with an antibody FV fragment.";

Proc. Natl. Acad. Sci. U.S.A. 94:10547-10553(1997).

RN [4]

RP COVALENT BOND.

RX MEDLINE=99268331; PubMed=10338009;

RA Buse G., Soulliane T., Dewor M., Meyer H.E., Blueggel M.;

RT "Evidence for a copper-coordinated histidine-tyrosine cross-link in

the active site of cytochrome oxidase.";

Protein Sci. 8:985-990(1999).

RN [5]

RP REVIEW.

RX MEDLINE=96391941; PubMed=8794157;

RA Ostermeier C., Iwata S., Michel H.;

RT "Cytochrome c oxidase";

Curr. Opin. Struct. Biol. 6:460-466(1996).

CC -/-

FUNCTION: Subunit I and II form the functional core of the enzyme

complex. Electrons originating in cytochrome c are transferred via

heme a and Cu(A) to the binuclear center formed by heme a3 and

Cu(B). This cytochrome c oxidase shows proton pump activity across

the membrane in addition to the electron transfer.

32 1738.5 58.0 527 2 Q6WAV9
33 1738.5 58.0 527 2 Q6YSM3
34 1738.5 58.0 527 2 AAP58355
35 1738.5 58.0 527 2 BAC98915
36 1738.5 58.0 527 2 Q7YFEL
37 1738.5 57.7 524 1 COX1_ORYSA
38 1738.5 57.7 524 2 Q7JAI6
39 1738.5 57.7 524 2 BAC19897
40 1738.5 57.7 528 1 COX1_MAIZE
41 1738.5 57.7 528 2 AAN33123
42 1738.5 57.7 528 2 AAR91047
43 1738.5 57.7 530 1 COX1_SORBI
44 1738.5 57.7 632 2 Q35812
45 1729.5 57.6 524 1 COX1_WHEAT

Q6WAV9 brassica ju
Q6YSM3 brassica na
AAP58355 brassica
BAC98915 brassica
Q7YFEL brassica ju
PL4578 cryza sativ
Q7JAI6 cryza sativ
BAC19897 oryza sat
P08742 zea mays (m
AAN33123 zea mays
AAR91047 zea mays
P05502 sorghum bic
Q35812 sorghum bic
P08741 triticum ae

ALIGNMENTS

FT	TURN	121	121	TURN	121
FT	HELIX	128	150	FT	HELIX
FT	STRAND	152	153	FT	STRAND
FT	TURN	154	156	FT	TURN
FT	STRAND	157	158	FT	STRAND
FT	TURN	163	166	FT	TURN
FT	TURN	168	170	FT	TURN
FT	HELIX	171	173	FT	HELIX
FT	HELIX	178	205	FT	HELIX
FT	TURN	210	211	FT	TURN
FT	HELIX	214	216	FT	HELIX
FT	HELIX	219	250	FT	HELIX
FT	TURN	255	256	FT	TURN
FT	HELIX	258	260	FT	HELIX
FT	TURN	261	261	FT	TURN
FT	HELIX	284	298	FT	HELIX
FT	TURN	299	299	FT	TURN
FT	TURN	303	304	FT	TURN
FT	HELIX	305	318	FT	HELIX
FT	TURN	319	320	FT	TURN
FT	HELIX	323	326	FT	HELIX
FT	HELIX	328	330	FT	HELIX
FT	TURN	331	331	FT	TURN
FT	HELIX	334	346	FT	HELIX
FT	TURN	347	347	FT	TURN
FT	HELIX	348	363	FT	HELIX
FT	STRAND	367	367	FT	STRAND
FT	HELIX	371	394	FT	HELIX
FT	HELIX	396	402	FT	HELIX
FT	TURN	403	404	FT	TURN
FT	STRAND	405	405	FT	STRAND
FT	HELIX	406	416	FT	HELIX
FT	TURN	417	419	FT	TURN
FT	HELIX	420	436	FT	HELIX
FT	STRAND	438	439	FT	STRAND
FT	HELIX	442	460	FT	HELIX
FT	HELIX	462	488	FT	HELIX
FT	TURN	469	470	FT	TURN
FT	STRAND	472	472	FT	STRAND
FT	HELIX	480	482	FT	HELIX
FT	HELIX	483	513	FT	HELIX
FT	STRAND	515	516	FT	STRAND
FT	TURN	525	526	FT	TURN
FT	HELIX	530	532	FT	HELIX
FT	TURN	533	533	FT	TURN
SEQ	SEQUENCE	558 AA;	62439 MW;	A8402453C0C0339E	CRC64;
Query Match 84.6%; Score 2538.5; DB 1; Length 558;					
Best Local Similarity 81.0%; Pred. No. 4.5e-156;					
Matches 452; Conservative 51; Mismatches 54; Indels 1; Gaps 1;					
Qy	1	MADAAIHGH-DHHEKQGFTRFMFSTNHXKDIGLLYLVAAGVGVGFISVLFVTVMRLRLMDP	59		
Db	1	MADAAVHGHDHDTGTFTRFMFSTNHXKDIGLLYLTAGIVGLISVCFVTVMRLMDLQHP	60		
Qy	60	GVQYMCLEGARLIADASQCTANGHLNWNMTYHGILMMFFVVGIPALFGFGNYLMPLQI	119		
Db	61	GVQYMCLEGARLIADASACTPNGHLNWNMTYHGVLMMFFVVIIPALFGFGNYFWPLHI	120		
Qy	120	GAPDMAFPRKNNLSFWLFTAGTAMGVASIFAPCGDQQLGSGGVNLYPPLSTREAGYSMD	179		
Db	121	GAPDMAFPRKNNLSYMYVCGVALGVASLIAPGNDQMGSGGVNLYPPLSTTTEAGYSMD	180		
Qy	180	LAIFAVHLSGASSIMCAINMTTFLNRPAGMTLHKVPLFSWSIFITAWLILLALPVLG	239		
Db	181	LAIFAVHVSASSILGAINIITFLNRPAGMTLFXVPLFANSVFFITAWLILLALSPLVLAG	240		
Qy	240	AITMLLTDNRNFGTTFPNAGGDPILYQHILMFPFGHPEVYIILPGFGIISHVVSFSKK	299		
Db	241	AITMLLMDNRNFGTQTFDPAGGDPVLYQHILMFPFGHPEVYIILPGFGIISHVISTFAKK	300		
Qy	300	PVFGYLPMYAYVAIGVLGVVWAHHMTVGMSLTQOSYFMLATMVAIVPTGIKIFSIA	359		

Db 301 PFGYLPVLAIAIGILGVVVAHMYTAGMSITQQAIFYMLATMTIAVPTGKIVFWSIA 360
 QY 360 TMWGSVEFKSPMLWAFGMFLFTVGGVTGIVLAQAGLDRAHDYDYYVVAHFHYVMSLGA 419
 Db 361 TMWGSIEFKPMLWAFGLFLFTVGGVTGVLSQAPLDLYDYYDYYVVAHFHYVMSLGA 420
 QY 420 IPATAGIYFMPKSGRAFPPEWAALHFWTFFIGANVTFFPQHLGRQGMPPRYIDYPE 479
 Db 421 VFGIPAGYVYVIGKSGRQYPEWAGQLHFWMMFIGNLFFPQHLGRQGMPPRYIDYPV 480
 QY 480 AFALWKNVSSYGAFLAFASLFFFTVIVYVTLVAGRRTRPNPWEADTLEWTLSPPPA 539
 Db 481 EFAYWNNISSGAYISFASLFFFTGIVYVTLVAGKRVNVVYVWNEHADTLEWTLSP 540
 QY 540 HTFETLPKRSWDKHPSH 557
 Db 541 HTFETLPKREDWDRAHAH 558

RESULT 2

COX1 RHOSH
 ID COX1 RHOSH STANDARD; PRT; 566 AA.
 AC P33517;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1) (Cytochrome a3 subunit 1).
 GN Name=ctaD;
 OS Rhodobacter sphaeroides (Rhodospseudomonas sphaeroides).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
 OC Rhodobacteraceae; Rhodobacter.
 OX NCBI_taxID=1063;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GA;
 RX MEDLINE=92204019; PubMed=1313140;
 RA Shapleigh J.P., Gennis R.B.;
 RT "Cloning, sequencing and deletion from the chromosome of the gene encoding subunit I of the a3-type cytochrome c oxidase of Rhodobacter sphaeroides.";
 RL Mol. Microbiol. 6:635-642(1992).
 RN [2]
 RP REVISIONS TO 436-439 AND 518-521.
 RA Shapleigh J.P., Gennis R.B.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Cytochrome c oxidase is the component of the respiratory chain that catalyzes the reduction of oxygen to water. Subunits 1-3 form the functional core of the enzyme complex. Co I is the catalytic subunit of the enzyme. Electrons originating in cytochrome c are transferred via the copper A center of subunit 2 and heme a of subunit 1 to the bimetallic center formed by heme a3 and copper B. This cytochrome c oxidase shows proton pump activity across the membrane in addition to the electron transfer.
 CC -!- CATALYTIC ACTIVITY: 4 ferrocytochrome c + O(2) = 4 ferricytochrome c + 2 H(2)O.
 CC -!- COFACTOR: Two heme groups and copper B.
 CC -!- PATHWAY: Respiratory chain; terminal step.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: Belongs to the heme-copper respiratory oxidase family.

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; X62645; CAA44514.2; -.
 DR PDB; 1M56; X-ray; A/G=1-566.
 DR PDB; 1M57; X-ray; A/G=1-566.
 DR InterPro; IPR000883; COX1.

DR Pfam; PF00115; COX1; 1.
 DR PRINTS; PR0165; CYCOXIDASEI.
 DR PROSITE; PS0855; COX1; 1.
 DR PROSITE; PS00077; COX1; CUB; 1.
 KW 3D-structure; Copper; Heme; Hydrogen ion transport; Oxidoreductase;
 KW Respiratory chain; Transmembrane.
 FT METAL 102 102 Iron (heme A axial ligand) (Probable).
 FT METAL 284 284 Copper B (Probable).
 FT METAL 288 288 Copper B (Probable).
 FT METAL 333 333 Copper B (Probable).
 FT METAL 334 334 Copper B (Probable).
 FT METAL 419 419 Iron (heme A3 axial ligand) (Probable).
 FT METAL 421 421 Iron (heme A axial ligand) (Probable).
 FT METAL 284 288 1'-histidyl-3'-tyrosine (His-Tyr) (By similarity).
 SQ SEQUENCE 566 AA; 63147 MW; 65A74BCCSC550B0 CRC64;
 Query Match 83.1%; Score 2493.5; DB 1; Length 566;
 Best Local Similarity 79.4%; Pred. No. 3.7e-153;
 Matches 450; Conservative 49; Mismatches 57; Indels 11; Gaps 2;
 QY 1 MADAALHGHDEHKEQGFTRFMSTNHKDIGLLYLVAAGVGFISVLFTVYRLEMDPG 60
 Db 1 MADAALHGHDEHKEQGFTRFMSTNHKDIGLLYLFTGGLVGLISVAFTVYRLEMDPG 59
 QY 61 VQWMLLEGAR-----LIADASQCTCTANGHLWNVVTVHGIILMFFVGPALFGGF 110
 Db 60 VQWMLLEGAR-----LIADASQCTCTANGHLWNVVTVHGIILMFFVGPALFGGF 119
 QY 111 GNYLMLPQIGAPDMAFPRMNNLSFWLFIAGTAGVASLFPAGDGGQGLSGVGNLYPPLS 170
 Db 120 GNYFMLPHIGAPDMAFPRMNNLSYWLIVAGTSLAVASLFPAGGNGQGLSGVGNLYPPLS 179
 QY 171 TREAGYSMDLAIPAVHLSGASSIMGAINMTTFLNMRAPGMTLHKYPLFSWSIFITAWLI 230
 Db 180 TSESGYSTDLAIPAVHLSGASSILGAINMTTFLNMRAPGMTLHKYPLFSWSIFITAWLI 239
 QY 231 LLALPVLAGAITMLLTDNRNFTFFNPAGGDDPILYQHILWFFGHDEPVIILPFGGIIS 290
 Db 240 LLALPVLAGAITMLLTDNRNFTFFNPAGGDDPILYQHILWFFGHDEPVIILPFGGIIS 299
 QY 291 HVSTFSKPKPVFGYLPNVYAMVAIGVLGVVVAHMYTVGMSLTQOSYFMLATMTIAVPT 350
 Db 300 HVIATFAKPKPIFGYLPNVYAMVAIGVLGVVVAHMYTVGMSLTQOSYFMLATMTIAVPT 359
 QY 351 GIKIFSWIATMWSGVVEFKSPMLWAFGMFLFTVGGVTGIVLAQAGLDRAHDYDYYVVAH 410
 Db 360 GIKIFSWIATMWSGVVEFKSPMLWAFGMFLFTVGGVTGIVLAQAGLDRAHDYDYYVVAH 419
 QY 411 FHYVMSLGAIFAIFAGIYFYMPKSGRAFPPEWAALHFWTFFIGANVTFFPQHLGRQGM 470
 Db 420 FHYVMSLGAIFAIFAGIYFYMPKSGRAFPPEWAALHFWTFFIGANVTFFPQHLGRQGM 479
 QY 471 PRYIDYPEAFALWKNVSSYGAFLAFASLFFFTVIVYVTLVAGRRTRPNPWEADTLE 530
 Db 480 PRYIDYPEAFALWKNVSSYGAFLAFASLFFFTVIVYVTLVAGRRTRPNPWEADTLE 539
 QY 531 WTLSPPPPAHTFETLPKRSWDKHPSH 557
 Db 540 WTLSPPPPAHTFETLPKRSWDKHPSH 558

RESULT 3

CX1A PARDE
 ID CX1A PARDE STANDARD; PRT; 554 AA.
 AC P08305;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Cytochrome c oxidase polypeptide I-alpha (EC 1.9.3.1) (Cytochrome A3 subunit 1-alpha).
 GN Name=ctaD; Synonyms=coi;
 OS Paracoccus denitrificans.


```

RA Capela D., Barloy-Hubler F., Couzy J., Bothe G., Ampe F., Batur J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Goffie I., Goffeau A., Kahn D., Kiss E., Lelaure V., Masny D.,
RA Pohl T., Portetelle D., Puhler A., Purnelle B., Rampsperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.,
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC -!- SIMILARITY: Belongs to the heme-copper respiratory oxidase family.
DR EMBL; AL591785; CAC45477.1; -.
DR HSP; P33517; IM56.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0005746; C: mitochondrial electron transport chain; IEA.
DR GO; GO:0004129; F: cytochrome-c oxidase activity; IEA.
DR GO; GO:0016491; F: oxidoreductase activity; IEA.
DR GO; GO:0006118; P: electron transport; IEA.
DR GO; GO:0006810; P: transport; IEA.
DR InterPro; IPR000883; COX1.
DR Pfam; PF00115; COX1; 1.
DR PROSITE; PS01165; CYCOXIDASE1.
DR Complete proteome; Electron transport; Heme; Oxidoreductase;
KW Respiratory chain; Transmembrane; Transport.
SQ SEQUENCE 562 AA; 62129 MW; 394144C173583612 CRC64;

Query Match 63.3%; Score 1898.5; DB 2; Length 562;
Best Local Similarity 63.2%; Pred. No. 1e-114;
Matches 348; Conservative 77; Mismatches 117; Indels 9; Gaps 7;

QY 3 DAAIGH-DH-HEKGGFTRFMSTNHKDIGLLYLVAAGVGVFISVLFTVMRLMDPG 60
DB 13 DHSDDHGHADHAKPTTFRRFLSTNHKDIGTLLYLFAIAGILGGTILSVFRAELQEPG 72

QY 61 VQVMCLG-ARLIADASQTCANG--HLWNVMVTYHGLMFFVGVIPALFGGFGNYLMLP 117
DB 73 IQ--IFHGLAQMVGFEQDAIDGGKHMFNVFTTAHALIMFFVMYPALIGGFANWVPI 130

QY 118 QIGAPDMAPPRNNLSFWLFIAGTAGNVAASLPAGCGDQLGSGVGVLYPPLSTR-RAGY 176
DB 131 MIGAPDMAPPRNNLSFWLFIAGTAGNVAASLPAGCGDQLGSGVGVLYPPLSTR-RAGY 190

QY 177 SMDLAI FAVHLSGASSIMGAINMITFLNMRAPGTLHKVPLFSWIFITAWLLILALPV 236
DB 191 AMDLAILGHTAGASSILGAINFITILNMRAPGTLHKVPLFAWSVLITAFLLLSLPV 250

QY 237 LAGAITMLLTDNRNFTTFNPAGGDDPILYQHILWFFGHPEVYIILPGFGIISHVWSTF 296
DB 251 LAGAITMLLTDNRNFTTFNPAGGDDPILYQHILWFFGHPEVYIILPGFGIISHVWSTF 310

QY 297 SKKPVGYLPMYVAIVAIGLVGVVVAHMYTVGMSLTQGSYVEMLATWIAVPTGIKIFS 356
DB 311 SKKPIFGYLGWYVAIVAIGLVGVVVAHMYTVGMSLTQGSYVEMLATWIAVPTGIKIFS 370

QY 357 WIATWGGSVFESKPMWAFGFMFLFTYGGVTGIVLAQAGLDRAVDHTYVVAHFHYVMS 416
DB 371 WIATWGGSVFESKPMWAFGFMFLFTYGGVTGIVLAQAGLDRAVDHTYVVAHFHYVMS 430

QY 417 LGAI FAI FAGIYFVMPKSGRAFPWAAKLHFWFFIGANVTFFPQHFLRGQHPRIYD 476
DB 431 LGAI FAI FAAYVYFVMPKSGRAFPWAAKLHFWFFIGANVTFFPQHFLRGQHPRIYD 490

QY 477 YPEAFALNKVSVYGAFLAFASFLFFIVFYVTLVAGRRTRPNPWGEGFADTLEWTLPS 536
DB 491 YPEAFALNKVSVYGAFLAFASFLFFIVFYVTLVAGRRTRPNPWGEGFADTLEWTLPS 549

QY 537 PPAHTFETLPK 547
DB 550 PPFHWEQQLPR 560

RESULT 5
Q8UHB4 PRELIMINARY; PRT; 572 AA.
ID Q8UHB4

```

```

AC Q8UHB4;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 25, Last annotation update)
DE Cytochrome-c oxidase chain I.
GN Names:coxA; OrderedLocusNames=Atu0768;
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=Dupont;
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P., Woo L.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendinning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).
CC -!- SIMILARITY: Belongs to the heme-copper respiratory oxidase family.
DR EMBL; AE009044; AAL41784.1; -.
DR HSP; P98002; IARI.
DR PUR; AB2671; AB2671.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0005746; C: mitochondrial electron transport chain; IEA.
DR GO; GO:0004129; F: cytochrome-c oxidase activity; IEA.
DR GO; GO:0016491; F: oxidoreductase activity; IEA.
DR GO; GO:0006118; P: electron transport; IEA.
DR GO; GO:0006810; P: transport; IEA.
DR InterPro; IPR000883; COX1.
DR Pfam; PF00115; COX1; 1.
DR PROSITE; PS01165; CYCOXIDASE1.
DR Complete proteome; Electron transport; Heme; Oxidoreductase;
KW Respiratory chain; Transmembrane; Transport.
SQ SEQUENCE 572 AA; 63022 MW; 1634BCF6E6005A4F CRC64;

Query Match 63.2%; Score 1897.5; DB 2; Length 572;
Best Local Similarity 63.9%; Pred. No. 1.2e-114;
Matches 350; Conservative 73; Mismatches 112; Indels 13; Gaps 6;

QY 7 HGHDDHEKQGFTRFMSTNHKDIGLLYLVAAGVGVFISVLFTVMRLMDPGVYMCL 66
DB 29 HSHDHSKPGFFARWFLSTNHKDIGTLLYLFAIAGILGGTILSVFRAELQEPG 86

QY 67 EG-ARLIADASQTCANG--HLWNVMVTYHGLMFFVGVIPALFGGFGNYLPLQIGAPD 123
DB 87 HGLASMYVGEQDAIDGGKHMFNVFTTAHALIMFFVMYPALIGGFANWVPI 146

QY 124 MAPFRNNLSFWLFIAGTAGNVAASLPAGCGDQLGSGVGVLYPPLSTR-EAGYSMDLAI 182
DB 147 MAPFRNNLSFWLFIAGTAGNVAASLPAGCGDQLGSGVGVLYPPLSTR-EAGYSMDLAI 206

QY 183 FAVHLSGASSIMGAINMITFLNMRAPGTLHKVPLFSWIFITAWLLILALPVLGAGIT 242
DB 207 FSLHVAGASSILGAINFITILNMRAPGTLHKVPLFAWSVLITAFLLLSLPVLGAGIT 266

QY 243 MLLTDRNFTTFNPAGGDDPILYQHILWFFGHPEVYIILPGFGIISHVWSTF 302
DB 267 MLLTDRNFTTFNPAGGDDPILYQHILWFFGHPEVYIILPGFGIISHVWSTF 326

QY 303 GYLPMYVAIVAIGLVGVVVAHMYTVGMSLTQGSYVEMLATWIAVPTGIKIFSWIATMW 362
DB 327 GYLPMYVAIVAIGLVGVVVAHMYTVGMSLTQGSYVEMLATWIAVPTGIKIFSWIATMW 386

```

363 GGSVEFKSPMLWAFGPMFLFTVGGTGTGIVLAQAGLDRAHYDTYVVVAHYVMSLGAIFA 422
 387 GGSLSLTESTPMVAIGFIPLFTVGGTGTGIVLAQAGLDRLSDYVVVAHYVMSLGAIFA 446
 423 IFAGIYFVYKPSGSRAPFPAWAKLHFWTFIGANVTFFQHFGLGQGMRRYIDYDEAFA 482
 447 IFAGWYFVYKPSGSRAPFPAWAKLHFWTFIGANVTFFQHFGLGQGMRRYIDYDEAFA 506
 483 LNKVSSYGAFL---AFASGLFFVIVFVTLVAGRETRPNWGPADTLEWTLSPPPA 539
 507 GNNVSSYSYSISAVAGVIFLP---GVNEAFKRIAGNNWPGEGATTLEWQLSSPPY 562
 540 HTPETLPK 547
 563 HOWEQLPR 570

RESULT 6
 Q8G263 PRELIMINARY; PRT; 552 AA.
 AC Q8G263;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Cytochrome c oxidase, subunit I (EC 1.9.3.1).
 GN Names: coxA; OrderedLocusNames=BR0468;
 OS Brucella suis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 NCBI_TaxID=29461;
 [1]
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=4330 / Biovar 1;
 RX MEDLINE=2224741; PubMed=12271122; DOI=10.1073/pnas.1923190989;
 RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
 Read T.D., Dodson R.J., Umayam L.A., Brinkac L.M., Bearan M.J.,
 Daugherty S.C., DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
 Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J.A.,
 Van Aken S.E., Riedmuller S., Tettelin H., Gill S.R., White O.,
 Salzberg S.B., Hoover D.L., Linder L.E., Hailing S.M., Boyle S.M.,
 Fraser C.M.;
 RA "The Brucella suis genome reveals fundamental similarities between
 RT animal and plant pathogens and symbionts.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
 CC -1- SIMILARITY: Belongs to the heme-copper respiratory oxidase family.
 DR EMBL; AE014356; AN29411.1; -;
 DR HSSP; P33517; IM56.
 DR TIGR; BR0468; -;
 DR GO; GO:0016021; C: integral to membrane; IEA.
 DR GO; GO:0005746; C: mitochondrial electron transport chain; IEA.
 DR GO; GO:0004129; F: cytochrome-c oxidase activity; IEA.
 DR GO; GO:0016491; F: oxidoreductase activity; IEA.
 DR GO; GO:0006118; P: electron transport; IEA.
 DR GO; GO:0006810; P: transport; IEA.
 DR InterPro; IPR000883; COXI.
 DR Pfam; PF00115; COXI; 1.
 DR PRINTS; PR01165; CYCOXIDASEI.
 DR PROSITE; PS00077; COXI; 1.
 DR Complete proteome; Electron transport; Heme; Oxidoreductase;
 KW Respiratory chain; Transmembrane; Transport;
 SQ SEQUENCE 552 AA; 60723 MW; 072B846C47D628 CRC64;

Query Match 63.0%; Score 1891.5; DB 2; Length 552;
 Best Local Similarity 63.0%; Pred. No. 2.9e-114;
 Matches 351; Conservative 76; Mismatches 113; Indels 17; Gaps 8;

1 MADAAATH---GHDHHEKQGFETFWFSTNKHDTGLLYVAAGVGFISLVFVYMLELM 57
 1 MAGTAAHGAHGDHDKPHG-WRVWVYSTNKHDTGLLYLIPAIAGIIGALSINRAELQ 59
 58 DPGVQVYKLEG-ARLI--ADASQCTCTANGHLMNVMVYTHGILMFFVGPALFGGFGNLY 114
 60 EPGIQ--IFHGLAQMVGVGEGDAALDAGKMFNVFTSAHALVMIFVMVFPALIGGFANWM 117

115 MPLOIGAPDMAPFRMNNLSFWLFIAGTAMGVSALFAPGGDQGLSGVGWVLYPPLSTR-E 173
 118 VPLMIGAPDMAPFRMNNLSFWLFIAGTAMGVSALFAPGGDQGLSGVGWVLYPPLSTR-E 177
 174 AGYSMDLAI FAVHLSGASSIMGAINMITTFLNRAAGMTLHKVPLFSWSIFITAMILLIA 233
 178 PGPAVDFAILAIHISGASSILGAINFITITLNNRAPGMTLHKVPLFAVAVLGFATLILL 237
 234 LPVLAGATMTLLDRNFGTTFNPNAGGGDPIYLQHLWFFGHPEVYIIILPGFGIISHVV 293
 238 LPVLAGATMTLLDRNFGTTFNPNAGGGDPIYLQHLWFFGHPEVYIIILPGFGIISHVV 297
 294 STFSKPEVFGVLPVYANVAIGVLGVVVAHMYTVMGSLTQOSYFMLATMVIATVGTGK 353
 298 STFSKPEVFGVLPVYANVAIGVLGVVVAHMYTVMGSLTQOSYFMLATMVIATVGTGK 357
 354 IFSWIATMWSGVFEKSPMLWAFGPMFLFTVGGTGTGIVLAQAGLDRAHYDTYVVVAHY 413
 358 IFSWIATMWSGVFEKSPMLWAFGPMFLFTVGGTGTGIVLAQAGLDRAHYDTYVVVAHY 417
 414 VMSLGAIFAIFAGIYFVYKPSGSRAPFPAWAKLHFWTFIGANVTFFQHFGLGQGMRR 473
 418 VLSLGAIFAIFAGIYFVYKPSGSRAPFPAWAKLHFWTFIGANVTFFQHFGLGQGMRR 477
 474 YIDYPEAFALWNVSSYGAFLAFASFLFFVIVFVTL---VAGRRETRPNWGPADTLE 530
 478 YIDYPEAFALWNVSSYGAFLAFASFLFFVIVFVTL---VAGRRETRPNWGPADTLE 533
 531 WTLSPPPPAHTFETLPK 547
 534 WQLSSPPPPHQBOLPR 550

RESULT 7
 Q8YFQ5 PRELIMINARY; PRT; 552 AA.
 AC Q8YFQ5;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-MAR-2004 (Tremblrel. 25, Last annotation update)
 DE CYTOCHROME C OXIDASE POLYPEPTIDE I (EC 1.9.3.1).
 GN OrderedLocusNames=BME11465;
 OS Brucella melitensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 NCBI_TaxID=29459;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=16M / ATCC 23456 / Biotype 1;
 RX MEDLINE=20020109; PubMed=11756688; DOI=10.1073/pnas.221575398;
 RA DelVecchio V.G., Kapral V., Redkar R.J., Patra G., Mijer C., Los T.,
 Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
 Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Gotsman E.,
 Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
 Haselkorn R., Kyrides N.C., Overbeek R.;
 RA "The genome sequence of the facultative intracellular pathogen
 RT Brucella melitensis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
 CC -1- SIMILARITY: Belongs to the heme-copper respiratory oxidase family.
 DR EMBL; AE009593; AAL52646.1; -;
 DR FIR; AC3435; AC3435.
 DR HSSP; P33517; IM56.
 DR GO; GO:0016021; C: integral to membrane; IEA.
 DR GO; GO:0005746; C: mitochondrial electron transport chain; IEA.
 DR GO; GO:0004129; F: cytochrome-c oxidase activity; IEA.
 DR GO; GO:0016491; F: oxidoreductase activity; IEA.
 DR GO; GO:0006118; P: electron transport; IEA.
 DR GO; GO:0006810; P: transport; IEA.
 DR InterPro; IPR000883; COXI.
 DR Pfam; PF00115; COXI; 1.
 DR PRINTS; PR01165; CYCOXIDASEI.
 DR PROSITE; PS00077; COXI; 1.
 DR Complete proteome; Electron transport; Heme; Oxidoreductase;
 KW Respiratory chain; Transmembrane; Transport;
 SQ SEQUENCE 552 AA; 60723 MW; 072B846C47D628 CRC64;

KW Complete proteome; Electron transport; Heme; Oxidoreductase;
 KW Respiratory chain; Transmembrane; Transport; CRC64;
 SQ SEQUENCE 552 AA; 60768 MW; 022505D26753E769 CRC64;

Query Match 63.0%; Score 1890.5; DB 2; Length 552;
 Best Local Similarity 62.8%; Pred. No. 3.4e-114;
 Matches 350; Conservative 77; Mismatches 113; Indels 17; Gaps 8;

Qy 1 MADAAIH---CHDHEKQGFTRFMSTNKHDKIGLLVLAAGVGVFTSVLFTVYMRLELM 57
 Db 1 MAGTAAHEHGADHDKPHG-WVRVWYTNKHDKIGTLTLYLFAIAGTGGGALSAMRAELQ 59

Qy 58 DPGVQVMCLEG-ARLI--ADASQTCATANGHLNWNVMTYHGTILMFFVIGIPALFGGFGNYL 114
 Db 60 EPGIQ--IFHGLAQVWYVEGDAALDACKMNFVFTSAHALVWFFVWMPALIGGPANWM 117

Qy 115 MPLOIGAPDMAPPRNNLSFWLFTAGTAGMVASLFAFGGQDQGLSGVGWVLYPPLSTR-E 173
 Db 118 VPLMIGAPDMAPPRMNNISFWLLPPALTLISLNFPGPAGGPGGGWTLYPFPSTSGQ 177

Qy 174 AGYSMDLAIFAVHLSGASSIMGAINMITFLNMRAPGNTLHKVPLFSWIFITAWLILLA 233
 Db 178 PGPAVDFAILAHISGASSIIGAINFITILNMRAPGNTLHKMPLFAWAVLTAFLILLS 237

Qy 234 LPVLAGATMLLDRNFOTTFENPAGGDPILYQHILWFFGHPEVYIIILPGFGIISHV 293
 Db 238 LPVLAGGITMLLDRNFOTTFAPDGGDPILYQHLEWFFGHPEVYIILOPGFGIVSHV 297

Qy 294 STESKPVFGVLYPMVYAVTGLGVVVAHMTVGMSTLQCSYFVMTATWIAVPTGIK 353
 Db 298 STESKRPFGVLYPMVYAVTGLGVVVAHMTVGMSTLQCSYFVMTATWIAVPTGIK 357

Qy 354 IFSWIATWGGSVFESKPMWAFGMFLFTVGGVGTIVLAQAGLDRAVHDYTYVVAHFY 413
 Db 358 IFSWIATWGGSVFESKPMWAFGMFLFTVGGVGTIVLAQAGLDRAVHDYTYVVAHFY 417

Qy 414 VMSLCAIFAFAIGIYFPMKSGRAFPWAAKLHPWTFPGANVTFPQHLGQGMPPR 473
 Db 418 VLSLGAFAIFAFAIGIYFPMKSGRAFPWAAKLHPWTFPGANVTFPQHLGQGMPPR 477

Qy 474 YIDYEAALNKKVSSYGAFLAFASFLFFIVFYVTL---VAGRETRPNPBGFAATLE 530
 Db 478 YIDYEDAVAGNMWISSYGSYISG----FAVLFLYVFEAKKEAGANPGDGATLE 533

Qy 531 WTLPSPPPAHTFETLPK 547
 Db 534 WOLSSPPPPHQQEQLPR 550

RESULT 8
 COX1 RHILE STANDARD; PRT; 538 AA.
 AC Q08855;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1) (Cytochrome A3 subunit 1).
 GN Name=ctaD; Synonyms=cox4;
 OS Rhizobium leguminosarum.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
 OX NCBI_TaxID=384;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94161496; PubMed=8117073;
 RA Gabel C., Bittiger M.A., Maier R.J.;
 RT "Cytochrome a3 gene regulation in members of the family Rhizobiaceae: comparison of copper and oxygen effects in Bradyrhizobium japonicum and Rhizobium tropici";
 RT and Rhizobium tropici";
 RL Appl. Environ. Microbiol. 60:141-148 (1994).
 CC -!- FUNCTION: Cytochrome c oxidase is the component of the respiratory chain that catalyzes the reduction of oxygen to water. Subunits 1-

3 form the functional core of the enzyme complex. CO I is the catalytic subunit of the enzyme. Electrons originating in cytochrome c are transferred via the copper A center of subunit 2 and heme A of subunit 1 to the bimetallic center formed by heme A3 and copper B.
 -!- CATALYTIC ACTIVITY: 4 ferrocytochrome c + O(2) = 4 ferricytochrome c + 2 H(2)O.
 -!- PATHWAY: Respiratory chain; terminal step.
 -!- SUBCELLULAR LOCATION: Integral membrane protein.
 -!- DEVELOPMENTAL STAGE: Free in soil (not as bacteroid).
 -!- SIMILARITY: Belongs to the heme-copper respiratory oxidase family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sb.ch).

EMBL; X74341; CAAS2388.1; -.
 PIR; S36424; S36424.
 HSP; P98002; IQLE.
 Inter-Pro; IPR000883; COX1.
 Pfam; PF00115; COX1; 1.
 PRINTS; PR01165; CYCOXIDASE1.
 PROSITE; PS00855; COX1; 1.
 PROSITE; PS00077; COX1; CUB; 1.
 DR Copper; Heme; Oxidoreductase; Respiratory chain; Transmembrane.
 FT TRANSMEM 37 57 Potential.
 FT TRANSMEM 84 104 Potential.
 FT TRANSMEM 121 141 Potential.
 FT TRANSMEM 170 190 Potential.
 FT TRANSMEM 208 228 Potential.
 FT TRANSMEM 259 279 Potential.
 FT TRANSMEM 291 311 Potential.
 FT TRANSMEM 329 349 Potential.
 FT TRANSMEM 362 382 Potential.
 FT TRANSMEM 401 421 Potential.
 FT TRANSMEM 438 458 Potential.
 FT TRANSMEM 476 496 Potential.
 FT METAL 82 82 Iron (heme A axial ligand) (Probable).
 FT METAL 265 265 Copper B (Probable).
 FT METAL 269 269 Copper B (Probable).
 FT METAL 314 314 Copper B (Probable).
 FT METAL 315 315 Copper B (Probable).
 FT METAL 400 400 Iron (heme A3 axial ligand) (Probable).
 FT METAL 402 402 Iron (heme A axial ligand) (Probable).
 FT METAL 402 402 1'-histidyl-3'-tyrosine (His-Tyr) (By similarity).
 FT CROSSLNK 265 269
 SQ SEQUENCE 538 AA; 59133 MW; 740CBFE6C902D9BF CRC64;

Query Match 62.5%; Score 1873.5; DB 1; Length 538;
 Best Local Similarity 64.2%; Pred. No. 4.1e-113;
 Matches 348; Conservative 63; Mismatches 110; Indels 21; Gaps 4;

Qy 7 HGDHHEKQGFTRFMSTNKHDKIGLLVLAAGVGVFTSVLFTVYMRLEMDGVQVMCL 66
 Db 15 HGDHHAHP---TCWRRTNKHDKIGTLTLYLFAIAGVIGAAISLAIRAEMLYPGVEYF-- 69

Qy 67 EGARLIADASQTCATANGHLNWNVMTYHGTILMFFVIGIPALFGGFGNYLMPLOIGAPDMAP 126
 Db 70 -----HNTHLYNVFTSHGVIMIFMWPMAMIGGFGNFWFLMTIGAPDMAP 115

Qy 127 PRMNLSEFLFIAGTAGVASLFAFGGQDQGLSGVGWVLYPPLSTR-EAGYSMDLAIFAV 185
 Db 116 PRMNLSEFLFIAGTAGVASLFAFGGQDQGLSGVGWVLYPPLSTR-EAGYSMDLAIFAV 175

Qy 186 HLGASSIMGAINMITFLNMRAPGNTLHKVPLFSWIFITAWLILLAIVLAGAITMLL 245
 Db 176 HIACASSILGAINFITILNMRAPGNTLHKMPLFAWAVLTAFLILLSLPLVLAGAITMLL 235

Qy 246 TDRNFGTTFENPAGGDPILYQHILWFFGHPEVYIIILPGFGIISHVSTFSKPVFGYL 305

Db 236 TDRNFGTTFAPSGGGDPLLYQHLLFFGPHGVVILILPGFGMISHIISTFSRKPVEGYI 295
 QY 306 PMYAMVAIGLVGVVWVAHMYTVGSLTQOSYFMLATVIAVPTGKIFPSMATWGGG 365
 Db 296 GMYAVAAIGLVGVVWVAHMYTVGSLTQOSYFMLATVIAVPTGKIFPSMATWGGG 355
 QY 366 VEFKSPMLAFGEMFLFTVGGVGTGIVLAQAGLDRAYHDYVYVVAHMYTVGSLTQOSYFMLATVIAVPTGKIFPSMATWGGG 425
 Db 356 IEZATPMLALAFELFTVGGVGTGIVLAQAGLDRAYHDYVYVVAHMYTVGSLTQOSYFMLATVIAVPTGKIFPSMATWGGG 415
 QY 426 GIFYVMPKSGRAPPWAALKHFWTFEIGANVTFFPHQHLGROGMPRRVIDYDPAFALWN 485
 Db 416 GWYVFPKSGMYNETLAEHFWLTFEIGANVTFFPHQHLGROGMPRRVIDYDPAFAGWN 475
 QY 486 KVSYSYGAFLAFASFLFFVIVFVTVLAGRRTRPNPWPBGFADTLEWTLSPPPPAFTFETL 545
 Db 476 LVSSIGSYISGFSVLLFIYCVYDAFAKNVPVGDNPWAGATLEWTLSPPPVHEFVL 534
 QY 546 PK 547
 Db 535 PR 536

RESULT 9

Q9A300 PRELIMINARY; PRT; 552 AA.
 AC Q9A300; 2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Cytochrome c oxidase, subunit I.
 GN OrderedLocusNames=CC3406;
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
 OC Caulobacteraceae; Caulobacter.
 OX NCBI_TaxID=155892;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647; DOI=10.1073/pnas.061092998;
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 Kolonay J.F., Smit J., Craven M.B., Khouri H.M., Shetty J.,
 Berry K.J., Utterback T.R., Tran K., Wolf A.M., Vamathevan J.J.,
 Emolaeva M.D., White O., Salzberg S.L., Venter J.C., Shapiro L.,
 Fraser C.M.;
 RT "Complete genome sequence of Caulobacter crescentus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 CC -1- SIMILARITY: Belongs to the heme-copper respiratory oxidase family.
 DR EMBL: AB006001; AAK25368.1; -;
 DR PIR: D87671; D87671.
 DR HSSP: P00396; Z0CC.
 DR TIGR: CC3406; -;
 DR GO: GO:0016021; C:integral to membrane; IEA.
 DR GO: GO:0005746; C:mitochondrial electron transport chain; IEA.
 DR GO: GO:0004129; F:cytochrome-c oxidase activity; IEA.
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.
 DR GO: GO:0006118; P:electron transport; IEA.
 DR GO: GO:0006810; P:transport; IEA.
 DR InterPro: IPR000883; COX1.
 DR Pfam: PF00115; COX1; 1.
 DR PRINTS: PR01165; CYCOXIDASE1.
 DR PROSITE: PS00077; COX1; 1.
 DR Complete proteome; Electron transport; Heme; Oxidoreductase;
 KW Respiratory chain; Transmembrane; Transport.
 SQ SEQUENCE 552 AA; 61101 MW; E82191E54206AACS CRC64;

Query Match 61.6%; Score 1847.5; DB 2; Length 552;
 Best Local Similarity 61.6%; Pred. No. 2e-111;
 Matches 341; Conservative 70; Mismatches 136; Indels 7; Gaps 4;

QY 2 ADAATHGHDEHKQGFTRFRMSTNHKIDIGLLVLAAGVVGFTSVLFTVVRMLDMPGV 61
 Db 5 ADTHDHAHDDKPPFPARFWFSFNHDKIDIGLLVLAAGVVGFTSVLFTVVRMLDMPGV 64
 QY 62 QYMCLEGARLIADASQCTTANGHLWNVYTVHGLMFFVGIIPALFGGFGNYLMPLOIGA 121
 Db 65 QVFSDTGML-----AQMGIFKKGHCYNAAVTAHALIMIFFMVPDAMIGGFGNFWPIMIGA 120
 QY 122 PDMAFPRNNLSFWLFTAGTAMGVASLFAFGDQGLSGVGVLYPLST-REAGYSMDL 180
 Db 121 PDMAFPRNNLSFWLFTAGTAMGVASLFAFGDQGLSGVGVLYPLST-REAGYSMDL 179
 QY 181 AIFAVHLSGASSIMGAINMITTFNNRPAQVTLHKVPLFSWFSFITAWLILLALPVLAGA 240
 Db 180 AILSLHLAGASSILGAINFITTFNMEAPGTLHRMPLFVWSVLITAFLLLSLPVLAGA 239
 QY 241 ITMLLTDRNFGTTFNAGGDDPILYOHILWFGHBEVYIILPGGIIISHVSTSKKP 300
 Db 240 ITMLLTDRNFGTTFNAGGDDPILYOHILWFGHBEVYIILPGGIIISHVSTSKKP 299
 QY 301 VFGYLPVYAMVAIGLVGVVWVAHMYTVGSLTQOSYFMLATVIAVPTGKIFSWIAT 360
 Db 300 VFGYLPVYAMVAIGLVGVVWVAHMYTVGSLTQOSYFMLATVIAVPTGKIFSWIAT 359
 QY 361 MWGGSVEFKSPMLWAFQGMFLFTVGGVGTGIVLAQAGLDRAYHDYVYVVAHMYTVGSLTQOSYFMLATVIAVPTGKIFSWIAT 420
 Db 360 MWGGSVEFKSPMLWAFQGMFLFTVGGVGTGIVLAQAGLDRAYHDYVYVVAHMYTVGSLTQOSYFMLATVIAVPTGKIFSWIAT 419
 QY 421 FAIFAGIYFYMPKFSGRAFPWAALKHFWTFEIGANVTFFPHQHLGROGMPRRVIDYDPEA 480
 Db 420 FAIFAGIYFYMPKFSGRAFPWAALKHFWTFEIGANVTFFPHQHLGROGMPRRVIDYDPEA 479
 QY 481 PALMNKVSYSYGAFLAFASFLFFVIVFVTVLAGRRTRPNPWPBGFADTLEWTLSPPPPAH 540
 Db 480 FTLWNVYSTVGYMTIVGVGVFVLMVLEIAIR-RRKAEANPWPBGFADTLEWTLSPPPPAH 538
 QY 541 TFEETLPKESDWDKH 554
 Db 539 QFSEPPVVKGGDHH 552

RESULT 10
 Q985X1 PRELIMINARY; PRT; 550 AA.
 ID Q985X1
 AC Q985X1; 2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Cytochrome c oxidase subunit I.
 GN OrderedLocusNames=mlr7491;
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti.";
 RL DNA Res. 7:331-338(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082936; PubMed=11214974;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,

RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.,
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti (supplement).",
RL DNA Res. 7:381-406(2000).
CC -1- SIMILARITY: Belongs to the heme-copper respiratory oxidase family.
DR EMBL; AP003011; BAB53582.1; -;
DR HSP; P33317; I456;
DR GO; GO:0016021; C: integral to membrane; IEA.
DR GO; GO:0005746; C: mitochondrial electron transport chain; IEA.
DR GO; GO:0004129; F: cytochrome-c oxidase activity; IEA.
DR GO; GO:0016491; F: oxidoreductase activity; IEA.
DR GO; GO:0006118; P: electron transport; IEA.
DR GO; GO:0006810; P: transport; IEA.
DR InterPro; IPR000883; COX1.
DR Pfam; PF00115; COX1; 1.
DR PRINTS; PR01165; CYCOXIDASE1.
DR PROSITE; PS00077; COX1; 1.
DR KW Complete proteome; Electron transport; Heme; Oxidoreductase;
KW Respiratory chain; Transmembrane; Transport.
SQ SEQUENCE 550 AA; 60241 MW; B3477BFA1B1AE67D CRC64;

Query Match 61.6%; Score 1846.5; DB 2; Length 550;
Best Local Similarity 61.4%; Pred. No. 2.3e-111;
Matches 341; Conservative 79; Mismatches 120; Indels 15; Gaps 8;

QY 1 MADAAIH-GDHHEKQGFTRWFSTNKHDTGLLYLVAAVGVGIFSVLFTVYMRLELMDP 59
Db 1 MADAAHGHCHDKPHG-WVRVYSTNKHDTGLYLFAINAGILGALSVAIRVELOEP 59
QY 60 GVQYMCLEG-ARLIADASQCTTANG--HLWNVMYTHGILMMFFVGIPALFGGNGVIMP 116
Db 60 GIQ--IFSGLAQMYGVNGDAIDGKSMYNAFAHAALIMFFWMPALIGGFANMVP 117
QY 117 LOIGAPDMAPRMMNLSFWLFIAGTAMGVASLEAPGGDGLGSGVGWLYXPLSTR-EAG 175
Db 118 IMIGAPDMAPRMMNISFWLLPPAPILLTSTNFTVPSAPGAFVGCGGNTLYPPLTSGQP 177
QY 176 YSMDLAI FAVHLSGASSINGAINMITTFNMRAPGMLTKVPLFSWSIFITAWLILALP 235
Db 178 PAMDALILSIHAGASSILGAINFITTFNMRAPGMLTKVPLFSWSIFITAWLILALP 237
QY 236 VLAGAITMLLTDNRNFTTFPAGGDPILYQHILWFFGHPEVYIIILPGGIIISHVYST 295
Db 238 VLAGGITMLLTDNRNFTTFPAGGDPILYQHILWFFGHPEVYIIILPGGIIISHVYST 297
QY 296 FSKKPVFGLPMYAMVAIGVLGVVVAHMYTVGMSLTQOSYFMLATMWIAVPTGKIF 355
Db 298 FSKKPVFGLMAYANVAIGVPIVVAHMYTVGMSLTQOSYFMLATMWIAVPTGKIF 357
QY 356 SWIATMWGSGVEFKPMLWAFGMFLFTVGGVTGIVLAQGLDRAYHDTYVVAHFHYVM 415
Db 358 SWIATMWGSGISFKTPEMLWALGFILFTIGTGGTVGLANAGLDRSLDHYFVIAHFHYL 417
QY 416 SIGALFAIPAGYFYFMPKESGRAPPEWAAKLHFWTFIFIGANVTFFPQHFGLRGQMPERY 475
Db 418 SUGAVFAIPAGYFYFMPKESGRAPPEWAAKLHFWTFIFIGANVTFFPQHFGLRGQMPERY 477
QY 476 DYPEAFALWNKVSYSYCAFLAFASFLFFIVFYITLVAGRRETR---PNPWGEFADTLWE 532
Db 478 DYPEAFAGNYSYSYISAVG---VAIPLYGVFEAFQKRIAGANPWGEFADTLWEQ 533
QY 533 LPSPPPAHTFTPLPK 547
Db 534 LPSPPPFHOWEQLPK 548

RESULT 11
Q6NBJ6 PRELIMINARY; PRT; 541 AA.
ID Q6NBJ6
AC Q6NBJ6;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)

DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Cytochrome c oxidase subunit I
DE Name=COXA; OrderedAccession=RPA0832;
GN Rhodospseudomonas palustris.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
CC Bradyrhizobiaceae; Rhodospseudomonas.
NCBI_TaxID=1076;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=GA009 / ATCC BAA-98;
RX PubMed=14704707; DOI=10.1038/nbt933;
RA Larimer F.W., Chain P., Hauser L., Lamerdin J.E., Malfatti S., Do L.,
RA Land M.L., Pelletier D.A., Beatty J.T., Lang A.S., Tabita F.R.,
RA Gibson J.L., Hanson T.E., Bobst C., Torres y Torres J.L., Peres C.,
RA Harrison F.H., Gibson J., Harwood C.S.;
RT "Complete genome sequence of the metabolically versatile
RT photosynthetic bacterium Rhodospseudomonas palustris";
RL Nat. Biotechnol. 22:55-61(2004).
CC -1- SIMILARITY: Belongs to the heme-copper respiratory oxidase family.
DR EMBL; BX572595; CAE26276.1; -;
DR InterPro; IPR000883; COX1.
DR Pfam; PF00115; COX1; 1.
DR PRINTS; PR01165; CYCOXIDASE1.
DR PROSITE; PS00077; COX1; 1.
DR KW Complete proteome; Electron transport; Heme; Oxidoreductase;
KW Respiratory chain; Transmembrane; Transport.
SQ SEQUENCE 541 AA; 59662 MW; 4C827D1C780FC41B CRC64;

Query Match 61.4%; Score 1842; DB 2; Length 541;
Best Local Similarity 63.1%; Pred. No. 4.5e-111;
Matches 346; Conservative 68; Mismatches 106; Indels 28; Gaps 7;

QY 7 HGHDDH---EKQGFTRWFSTNKHDKIGLLYLVAAGVGIFSVLFTVYMRLELMDPGVQY 63
Db 13 HAHDHDAHAPTC-WRRVLYSTNKHDKIGTMYLVFAVMAGVIGASIAIRVELMFFGIQI 71
QY 64 MCEGARLIADASQCTTANGHLWNVMYTHGILMMFFVGIPALFGGNGVLMPLQIGAPD 123
Db 72 F-----HNTHYVFTSHGLNIFWVMPAMIGGFWNFVLMIGAPD 115
QY 124 MAPRMMNISFWLFIAGTAMGVASLEAPGGDGLGSGVGWLYVPLST-REAGYSMDLAI 182
Db 116 MAPRMMNISFWLLPAAFALLIISTFVEGEPSSNGVGAGTWYAPLSTTICHPGPAVDFAI 175
QY 183 FAVHLSGASSINGAINMITTFNMRAPGMLTKVPLFSWSIFITAWLILALPVLGAI 242
Db 176 LALHLGASSILGAINFITTFNMRAPGMLTKVPLFSWSLITAFLLLSLPLVLAGAI 235
QY 243 MLLTDNRNFTTFPAGGDPILYQHILWFFGHPEVYIIILPGGIIISHVYSTFSKPVF 302
Db 236 MLLTDNRNFTTFPAGGDPILYQHILWFFGHPEVYIIILPGGIIISHVYSTFSKPVF 295
QY 303 GYLPMYAMVAIGVLGVVVAHMYTVGMSLTQOSYFMLATMWIAVPTGKIFSWIATMW 362
Db 296 GYLGMAYAMVAIGVLGVVVAHMYTVGMSLTQOSYFVAATMWIAVPTGKIFSWIATMW 355
QY 363 GGSVEKSPMLWAFGMFLFTVGGVTGIVLAQGLDRAYHDTYVVAHFHYVMSLGAIFA 422
Db 356 GGSIEFTKPLWAVGFIPLFTVGGVTGIVLANAGVRLVQDTYVVAHFHYVMSLGAIFA 415
QY 423 IFAGIYFMPKESGRAPPEWAAKLHFWTFIFIGANVTFFPQHFGLRGQMPERYDYPEAFA 482
Db 416 IFAGIYFMPKESGRAPPEWAAKLHFWTFIFIGANVTFFPQHFGLRGQMPERYDYPEAFA 475
QY 483 LWNKVSYSYCAFLAFASFLFFIVFY--TLVAGRRE--TRNPWGEFADTLWEPLSPPPA 539
Db 476 GWNLVSSIGSYISG---FAVLVFLYGMTLAFIRKRAADNPWGAGATTLEWTLSSPPPF 531
QY 540 HTEPLPK 547
Db 532 HQFEVLQ 539

DB 344 KVFSGIATMGGSIETPKTLMPLFAIGFIFLFTVGGVTGVVLANAGVDRVMOETVYVIAHFH 403
 QY 413 YVMSLGAIFALFAGFYFPMKFSGRAPPEMAAKLHFWTFIFIGANVTFFPQHFLGRQGMPR 472
 DB 404 YVMGIAVFAIFSGWYHFFPMGSGWYNETIGKLFHFWMFIGNILFFPQHFLGLQGMWR 463
 QY 473 RYIDYPEAFALWNVSSVGAFLAFASPLFFIVFYVTLV----AGRETNRPNWGEFADTL 529
 DB 464 RMVDYDPAFAGWNETISSYGAFITG----FGVIIFLYGLVDFAFARKQAADNPWGEGATTL 519
 QY 530 EWTLPSPPPPAHTFETLPK 547
 DB 520 EWTLSPPPPFHQFSTLPK 537

RESULT 14
 COX1_BRAJA STANDARD; PRT; 541 AA.
 AC P31833;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1) (Cytochrome AA3 subunit 1).
 GN Name=ctaD; Synonyms=coxA; OrderedLocusNames=blr1171;
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bradyrhizobiaceae; Bradyrhizobium.
 OX NCBI_TaxID=375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LO;
 RX MEDLINE=91045095; PubMed=2172930;
 RA Gabel C., Maier R.J.;
 RT "Nucleotide sequence of the coxA gene encoding subunit I of cytochrome a3 of Bradyrhizobium japonicum.";
 RL Nucleic Acids Res. 18:6143-6143(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=USDA 110spc4;
 RX MEDLINE=91211625; PubMed=1965217;
 RA Bott M., Bolliger M., Hennecke H.;
 RT "Genetic analysis of the cytochrome c-a3 branch of the Bradyrhizobium japonicum respiratory chain.";
 RL Mol. Microbiol. 4:2147-2157(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=USDA 110spc4;
 RX MEDLINE=22484998; PubMed=12597275;
 RA Rossmann R., Loferer H., Rossi P., Hennecke H.;
 RT "Factors involved in biogenesis of active cytochrome a3 encoded by the coxAEGC gene cluster from Bradyrhizobium japonicum.";
 RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=USDA 110spc4;
 RX MEDLINE=91211625; PubMed=1965217;
 RA Bott M., Bolliger M., Hennecke H.;
 RT "Genetic analysis of the cytochrome c-a3 branch of the Bradyrhizobium japonicum respiratory chain.";
 RL Mol. Microbiol. 4:2147-2157(1990).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=USDA 110;
 RX MEDLINE=22484998; PubMed=12597275;
 RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T., Sasanoto S., Watanabe A., Idesawa K., Iriquchi M., Kawashima K., Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M., Tabata S.;
 RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110.";
 RL DNA Res. 9:189-197(2002).
 CC -1- FUNCTION: Cytochrome c oxidase is the component of the respiratory chain that catalyzes the reduction of oxygen to water. Subunits 1-3 form the functional core of the enzyme complex. CO I is the

CC catalytic subunit of the enzyme. Electrons originating in
 CC cytochrome c are transferred via the copper A center of subunit 2
 CC and heme A of subunit 1 to the binuclear center formed by heme A3
 CC and copper B.
 CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome
 CC c + 2 H(2)O.
 CC -1- PATHWAY: Respiratory chain; terminal step.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- DEVELOPMENTAL STAGE: Free in soil (not as bacteroid).
 CC -1- SIMILARITY: Belongs to the heme-copper respiratory oxidase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X54318; CAA38216.1; -;
 CC EMBL; X54800; CAA38570.1; -;
 CC EMBL; AJ242592; CAB56819.1; -;
 CC EMBL; U33883; AAF78815.1; -;
 CC EMBL; AP005339; BAC46436.1; -;
 CC PIR; S13076; ODZJ1.
 CC HSP; P00396; 20CC.
 CC InterPro: IPR000883; COX1.
 CC Pfam; PF00115; COX1.1.
 CC PRINTS; PR0165; CYCOXIDASEI.
 CC PROSITE; PSS0855; COX1.1.
 CC PROSITE; PS00077; COX1_CUB; 1.
 CC Complete proteome; Copper; Heme; Oxidoreductase; Respiratory chain;
 CC Transmembrane. 40 60 Potential.
 CC TRANSMEM 87 107 Potential.
 CC TRANSMEM 124 144 Potential.
 CC TRANSMEM 173 193 Potential.
 CC TRANSMEM 211 231 Potential.
 CC TRANSMEM 262 282 Potential.
 CC TRANSMEM 294 314 Potential.
 CC TRANSMEM 332 352 Potential.
 CC TRANSMEM 365 385 Potential.
 CC TRANSMEM 404 424 Potential.
 CC TRANSMEM 441 461 Potential.
 CC TRANSMEM 483 503 Potential.
 CC METAL 85 85 Iron (heme A axial ligand) (Probable).
 CC METAL 268 268 Copper B (probable).
 CC METAL 272 272 Copper B (probable).
 CC METAL 317 317 Copper B (probable).
 CC METAL 318 318 Copper B (probable).
 CC METAL 403 403 Iron (heme A3 axial ligand) (Probable).
 CC METAL 405 405 Iron (heme A axial ligand) (Probable).
 CC METAL 426 426 1'-histidyl-3'-tyrosine (His-Tyr) (By
 CC CROSSLINK 268 272 similarity).
 CC SEQUENCE 541 AA; 59265 MW; 08F9A69D0C90DB85 CRC64;
 CC
 CC Query Match 61.0%; Score 1829; DB 1; Length 541;
 CC Best Local Similarity 61.8%; Pred. No. 3,le-110;
 CC Matches 337; Conservative 72; Mismatches 112; Indels 24; Gaps 4;
 CC
 CC QY 7 HGHDEHEKQGFETRWMTNKHDKIGLLYLAAGVVGVSIVLFTVYMRLELMDPGVQVWML 66
 CC DB 15 HGHDEHAHPTGWRVYVSTNKHDKIGLLYLAAGVVGVSIVLFTVYMRLELMDPGVQVWML 72
 CC QY 67 EGARLIADASCTCTANGHLNVMVTYIGILMFFVGPIDALFGGNGNVMPLQIGAPDMAP 126
 CC DB 73 -----HETHYVNFVTSGLINIFFMVMPAMITGGNFGNFMVPLMIGAPDMAP 118
 CC QY 127 PRMNLSFWLFIAGTAMGVASLPAFGDGGQGLGSGVGVVLYPPLSTR-EAGYSMDLAFV 185
 CC DB 119 PRMNISFWLPAFGGLLMSFTVEGPGANGVAGTMYVPLSSSGHPGPAVDFAISL 178
 CC QY 186 HLGASSINGAINMTTFLNRPAPMTLHKVPLPFSWISFIFATWLLILLPVLGAITWLL 245

```

Db 179 HLAGASSILGAINFITTIFNRRAGMTLHKRNPFLVMSILVTVLELLLSLVLGAGATMLL 238
Qy 246 TDRNFGTTFNPGAGGDPILYQHILWFFGHPEVYIIILPGFGIISHVSVFSFKPKPVFGYL 305
Db 239 TDRNFGTTFNPGAGGDPVLFQHLWFFGHPEVYIIILPGFGIISHVSVFSFKPKPVFGYL 298
Qy 306 PMYAMVAIGVLGRVWHAHMYTVGMSLTQOSFEMLATWIAVPTGKIFSWATATWGS 365
Db 299 GMATAMVAIGGIVGFWHAHMYTVGMSLTQOSFEMLATWIAVPTGKIFSWATATWGS 358
Qy 366 VEFKSPMLWAFGFMFLFTVGGVGTGIVLAQAGLDRAVHDYVVAHMYVMSLGAIFAIFA 425
Db 359 IEFAPAMVAVGPIFLFTVGGVGTGIVLAQAGLDRAVHDYVVAHMYVMSLGAIFAIFA 418
Qy 426 GIYFYMFKFSGRAPPEAAKLHFWTFPIGANVTFPQHFILGROGMPRRYIDYPEAFALWN 485
Db 419 GWYWFPEKMTGYMYNETLAKAHFWTFPIGANVTFPQHFILGROGMPRRYIDYPEAFALWN 478
Qy 486 KVSSYGAFALAFASPLFPFIVFVTVLVRGRTETP---NPMGEFADTLEWTLSPPPAHTF 542
Db 479 LVSSVGSYISG---FGVLFLICVIDAFKVPAGNPNWAGATTLEWTLSPPPAHTF 534
Qy 543 ETLPK 547
Db 535 EVLPR 539

```

RESULT 15

```

Q36520 PRELIMINARY; PRT; 523 AA.
AC Q36520;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Cytochrome oxidase subunit 1.
GN Names-coxi;
OS Platymonas subcordiformis.
OC Mitochondrion.
OC Eukaryota; Viridiplantae; Chlorophyta; Prasinophyceae; Platymonas.
CX NCBI_TaxID=3161;
RN [1]
RP MEDLINE=96145517; PubMed=8555450;
RX Kessler U.; Zesche K.;
RT "Physical map and gene organization of the mitochondrial genome from
RT the unicellular green alga Platymonas (Tetraselmis) subcordiformis
RT (Prasinophyceae).";
RL Plant Mol. Biol. 29:1081-1086(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Kessler U.;
RL Thesis (1995), Institute for Plant Physiology, Justus Liebig
RL University.
CC -1- FUNCTION: Cytochrome c oxidase is the component of the respiratory
CC chain that catalyzes the reduction of oxygen to water. Subunits 1-
CC 3 form the functional core of the enzyme complex. CO I is the
CC catalytic subunit of the enzyme. Electrons originating in
CC cytochrome c are transferred via the copper A center of subunit 2
CC and heme A of subunit 1 to the binuclear center formed by heme A3
CC and copper B (By similarity).
CC -1- CATALYTIC ACTIVITY: 4 ferrioxochrome c + O(2) = 4 ferrioxochrome
CC c + 2 H(2)O.
CC -1- PATHWAY: Respiratory chain; terminal step.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane (By similarity).
CC -1- SIMILARITY: Belongs to the heme-copper respiratory oxidase family.
DR EMBL; Z47795; CAA87753.1; -.
DR PIR; S62706; S62706.
DR HSP; P00396; 20CC.
DR GO; GO:0019866; C:inner membrane; IEA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0005746; C:mitochondrial electron transport chain; IEA.

```

```

DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0004129; F:cytochrome-c oxidase activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR000983; COX1.
DR Pfam; PF00115; COX1; 1.
DR PRINTS; PRO1165; CYCOXIDASE1.
DR PROSITE; PS00077; COX1; 1.
DR PROSITE; PS00077; COX1; 1.
KW Copper; Electron transport; Heme; Inner membrane; Membrane;
KW Mitochondrion; Oxidoreductase; Respiratory chain; Transmembrane;
KW Transport.
SQ SEQUENCE 523 AA; 57669 MW; 3E5B17B6E749085C CRC64;

Query Match 59.1%; Score 1773; DB 2; Length 523;
Best Local Similarity 60.1%; Pred. No. 1.3e-106;
Matches 322; Conservative 82; Mismatches 104; Indels 28; Gaps 6;

Qy 17 FFTFMSTNKHDKIGLLYVAAGVVGFSIVLVYVMELEMDPGVQVMCLGARLIADAS 76
Db 4 PAQRALSTNKHDKIGLLYVAAGVVGFSIVLVYVMELEMDPGVQVMCLGARLIADAS 76
Qy 77 QTCTANGHLNMYVYTHGILMMFFVGIPALFGGFGNYLMLQIGAPDPAFPRMNNISFWL 136
Db 49 QVLGNHQLYNYVITAHAFIMFFMNPALIGFGNWFVIMIGAPDPAFPRMNNISFWL 108
Qy 137 FIATAMGVASLFAFGGQGLGSGVGVLYPELST--REAGYSMDLAIFAVHLSGASSIM 194
Db 109 LPPSLLLSLSALV-----EIGAGTGMTVYPPSSIASHSGASVDLAIFSLHSGASSIL 163
Qy 195 GAINMITTFLNMRAPGMTLHKVPLFSIITAWLILLALPVLGATITMLTDRNFCTTF 254
Db 164 GALNFITILNMRGPGMTLHKVPLFSIITAWLILLALPVLGATITMLTDRNFCTTF 223
Qy 255 FNPAGGDPILYQHILWFFGHPEVYIIILPGFGIISHVSVFSFKPKPVFGYLPMYAVAI 314
Db 224 FDPAGGDPILYQHILWFFGHPEVYIIILPGFGIISHVSVFSFKPKPVFGYLPMYAVAI 283
Qy 315 GVLGVVWHAHMYTVGMSLTQOSFEMLATWIAVPTGKIFSWATATWGSVEFKSPMLW 374
Db 284 GILGFIWHAHMYTVGMSLTQOSFEMLATWIAVPTGKIFSWATATWGSVEFKSPMLW 343
Qy 375 AFGMFLFTVGGVGTGIVLAQAGLDRAVHDYVVAHMYVMSLGAIFAIFAIFAGYFYMFKF 434
Db 344 AIGFLFTVGGVGTGIVLAQAGLDRAVHDYVVAHMYVMSLGAIFAIFAIFAGYFYMFKF 403
Qy 435 SGRAFPPEAAKLHFWTFPIGANVTFPQHFILGROGMPRRYIDYPEAFALWNKVSYSYGAEL 494
Db 404 CGLOYSEFLQIHFWLFFVGVNITFFPMHFLGISGMPPRRIPDIPDAFAGNVAVCVSGYL 463
Qy 495 AFASFLFFIVFVTVLVRGRTETP---GEFADTLEWTLSPPPAHTFPTLP 546
Db 464 SVLGALFFVYV-VYDLSGERCPKPNWETVPGTSA-TLEWTLSPPPAHTFPEVP 517

```

Search completed: October 18, 2004, 23:44:31
Job time : 201 secs